ASSESSMENT OF NON-VERBAL INTELLIGENCE IN SOUTH AFRICAN SCHOOLS: DO LANGUAGE AND GENDER BIAS PERFORMANCE ON THE RAVEN'S STANDARD PROGRESSIVE MATRICES?

Jolene Knowles (9802626A)

A Research Report submitted to the Faculty of Humanities, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Educational Psychology by Coursework and Research Report.

Johannesburg, November 2008

DECLARATION

I hereby declare that this research report is my own independent work, and has not been presented for any other degree at any other academic institution, or published in any form.

It is submitted in partial fulfilment of the requirements for the degree of Masters of Educational Psychology by Coursework and Research Report at the University of the Witwatersrand, Johannesburg.

____

Jolene Knowles November 2008
(9802626A)

ACKNOWLEDGEMENTS

I wish to extend sincere thanks and grateful acknowledgement to the following people:

Dr. Broom, supervisor, for always being readily available for consultation. Your support throughout this year has been incredible. Thank you so much for your academic guidance, personal wisdom and exceptional patience. I am deeply grateful to you for everything. The opportunity to work with you has been a privilege.

Ms. Nicky Israel, for always being available and willing to help. Thanks you for your constant assistance and extraordinary patience. Your knowledge and expertise were much appreciated.

Mr. Mike Greyling for you patience and perseverance when I needed statistical assistance. Without your insights this research report would not have been possible.

Mr Peter Fridjon for your willingness to discuss statistical and psychometric theory.

Deborah Catto for your constant support. We got through this together and for that I am truly grateful. Working with you has been a privilege and I treasure our professional relationship and friendship.

To my husband, Brad, without you none of this would have been possible. You have been a rock throughout this process. Thanks for all your encouragement, patience, support and constant love when I needed it most.

To my parents, and parents in law, you have been so encouraging and have accommodated my every need. You all are such a blessing to me!

And to the rest of the M. Ed group and the rest of my family… thanks for everything.

ABSTRACT

Test bias within the field of psychometrics is an issue of concern in the South African context. Bias refers to whether a test measures what it claims to measure across different groups. In South Africa, psychological testing has been associated with the oppressive and discriminatory practices of the Apartheid era, leading to many tests being banned and considered invalid and unfair (Foxcroft & Roodt, 2004). Research is required in South Africa to investigate the effects and functions of psychometric tests in the context of the country's history and diverse population groups. This research investigated whether the Raven's Standard Progressive Matrices (SPM) is a good measure of non-verbal intelligence for high school children in the South African context, or whether it is biased by language and gender among Grade 8 and Grade 9 learners. The subjects were between 13 and 15 years old and attended an English medium government high school in Johannesburg. The archival data was collected over a two year period and the SPM was administered to some subjects in both Grade 8 and 9. A 2-way ANOVA tested for an overall difference on the SPM scores between learners who spoke English as a first language (EFL) and English as an additional language (EAL), as well as the performance of Male and Female learners. There was a significant difference between the language groups in Grade 8 but not Grade 9. There were no significant gender differences. An item analysis tested for a language or gender difference on the item level. Certain test items showed a gender difference while others a language difference, but no significant trend was found across the test. A matched paired t-test for the group that received the test in both Grade 8 and 9 determined that there was a significant increase in scores with greatest gains for the Male and Female EAL learners.

TABLE OF CONTENTS

LIST OF TABLES:

LIST OF FIGURES

CHAPTER ONE: LITERATURE REVIEW

1. 1. Introduction

Over time, a body of theory and research has been developed regarding scientific measurement principles which have been applied to psychological assessment tests, to ensure such tests are reliable and valid (Foxcroft & Roodt, 2004). Many studies have been conducted to establish the reliability and validity of these tests. One such test is the Raven's Progressive Matrices (RPM), a psychometric test of non-verbal intelligence, which was originally standardised on a British population. The RPM measures a component of Spearman's 'g', the educative (non-verbal) ability, which is the ability to make sense of complex situations, squeeze meaning out of events, and perceive and think clearly (Raven, 1965, 1994). There are three forms of the Raven's Progressive Matrices. The easy level is called the Ravens Coloured Progressive Matrices (CPM), which was designed for children between the ages of 5 and 11, as well as elderly people. The average level of the test is known as the Raven's Standard Progressive Matrices (SPM) which was designed for the general population from ages 6 to 80. Finally, the difficult level of the test, the Advanced Progressive Matrices (RAPM), was designed for age 11 and upwards (Arendasy & Sommer, 2005; Raven, 2003).

The RPM is a widely used test of non-verbal intelligence (Raven, 2003). Non-verbal intelligence is the ability to analyse information and solve problems using visual reasoning. Non-verbal tasks may involve concrete or abstract ideas, internalized language-based reasoning, and internalized reasoning without language. Non-verbal intelligence tests have been used as an attempt to assess intellectual aptitude while removing language barriers. These have been useful in assessing children without speech or limited language ability (Logsdon, 2008). The RPM has been applied to psychology (Ackerman & Kanfer, 1993; Arthur, Barret and Dorespike, 1990) and basic cognitive processing research (Ackerman, 1988; Ackerman, 1990; Babcock, 1994). It has allowed for a substantial database of performance profiles and cross-cultural comparisons, and can be administered on an individual basis or in a group setting (Raven, Raven & Court, 1998). The RPM has also been deemed useful as it contains a large number of items, making it appropriate for detailed statistical analysis (Carpenter, Just & Snell, 1990; DeShon, Chan & Weissbein, 1995). There is also a large data base of performance profiles and norms for different populations, countries and cultures (Owen, 1992). This research investigated whether the SPM, specifically, is a good measure of non-verbal intelligence in the South African context. The test is considered culturally, linguistically and gender fair which makes it potentially appropriate for South Africa's population, which is culturally, linguistically and racially diverse (Jensen & Feuerstein, UD; Owen, 1992).

In South Africa many people were disadvantaged due to the Apartheid regime. The government passed legislation such as the Native Land Act of 1913 and the Group Areas Act of 1950, which severely compromised the well-being and by implication, the education and development of many non-white South Africans (Mayekiso & Tshemse, 2007). With the changes in policy since 1994, there has been a move, in South Africa, toward English being the medium of instruction in many schools. In South Africa, children with similar socio economic status often have major differences in language. A child for whom English is not their first language may experience educational and language difficulties, even when they have received primary education in English. The question that is posed is: if learning in an additional language affects one's school performance, does this translate into a lowered IQ score on a test that was standardised on an English speaking population. Language bias may be a factor with regards to the SPM as children from families that were previously disadvantaged and for whom English is an additional language, may not be familiar with the constructs of the test, which might negatively affect their performance in the test. Thus, language is an important variable to consider when investigating whether the test is biased toward certain children in South Africa.

Gender has also been widely researched in terms of performance on the SPM. Many studies have suggested that there are no differences, or very small differences, in scores between boys and girls (e.g. Colom & Abad, 2007; Court, 1983, Jensen, 1998, Mackintosh, 1998, Raven, 1939; Lynn & Irwing, 2004,), while other studies such as Abad, Colom, Rebollo & Escorial (2004), Blinkhorn (2005), Colom, Escorial & Rebollo (2004), Lynn et al. (2004), Mackintosh & Bennet (2005) and have suggested that there are gender differences on particular items of the RPM, with mean scores favouring males over females. This research investigated whether the gender bias, previously reported and disputed in research, applied to South African high school children.

This study was a parallel study to the one conducted by Israel (2006). She examined systematic differences in performance on the RAPM, on the basis of home language and gender in one hundred first-year university students. Israel (2006), reported a substantial language bias at item level for university students. She suggested that strong performance on the RAPM was correlated to good English comprehension, making the test items unequally difficult across language groups. Israel's findings raised doubt about the linguistic and cultural fairness of the RAPM as a non-verbal test of intelligence.

The current study explored whether there was any evidence of bias in the SPM on the basis of language and gender. Bias was explored by an investigation of differences in performance between children for whom English is their first language (EFL) and children for whom English is an additional language (EAL) in the South African context. Bias was also investigated with regard to differences in performance on the basis of gender.

1.2. Ravens Progressive Matrices: Theoretical Background

The Raven's Progressive Matrices (RPM) were developed by Dr John C. Raven in 1938. The term "Progressive" refers to how the problems within each of the test variants becomes progressively more difficult (Gregory, 2007). This section examines the theoretical background concerning the RPM. The literature concerning what the test measures with regard to intelligence is also explored.

The first intelligence test was developed by Alfred Binet in the early 1900s for the purpose of determining which children could and could not attend the regular school programme in Great Britain. Binet described intelligence in terms of human judgment, initiative, and adaptation to circumstances with emphasis on attention and memory (Ashman & Conwah, 1991). It was not long after Binet's work that the notion of intelligence quotient (IQ) was defined as the mental age divided by chronological age (Gould, 1981, cited in Ashman et el, 1991).

The RPM measures non-verbal reasoning which is associated with fluid intelligence and general intellectual functioning or *g*. The more *g* loaded a test, the better it is able to predict academic achievement, creativity, career potential and job performance (Kuncel, Hezlett & Ones, 2004; Rohde & Thompson, 2007). The RPM and Vocabulary tests were based on observations and theories formulated by Spearman at the turn of the last century (Gregory, 2007). Spearman construed that general ability or *g* is the highest common factor in all measures of cognitive ability, and to some scholars this has been accepted as an empirical fact (Jensen, 1993).

Charles Spearman found that school children's grades across learning subject areas were positively correlated. He termed these correlations as general intelligence or *g*. He developed a two factor model explaining all variations in intelligence test scores. The first factor was specific to an individual mental task, making an individual more skilled at one cognitive task compared to another cognitive task. The second was *g*, a general factor that governs performance on all cognitive tasks. Tests of cognitive ability derive most of their validity from the extent to which they measure *g*. A test was said to be *g*-loaded if quantifiable measures of performance on a number of tasks highly correlate. Tests were seen to be more reliable and valid if they are as *g*-loaded as possible. According to Raven, Raven and Court (2000, p. 34), the "progressive matrices was described as one of the purest and best measures of *g* or general intellectual functioning". Some scientists argued that there is no single measure of intelligence, and general intelligence should be exposed as not well correlated within the various intellectual capacities. However, the notion of *g* has not been dismissed and is still employed as a valid evaluation of human mental ability (Raven, 1994).

The RPM is based on figural stimuli and is made up of multiple choice tests of abstract reasoning (Gregory, 2007). Raven, Raven & Court (1998) stated that the RPM is a measure of educative ability and should not be described as a measure of "general intelligence", "ability", or "problem-solving ability" (p. 7), but should rather be considered a measure of a range of abilities that are "built one on top of the other" (p. 7). He added that it is not generally possible to solve the more difficult problems without the ability to solve the easier ones. He argued that the abilities that are required to solve more difficult problems are quantitatively different from the abilities required to solve easier items, these "apparently different abilities shade imperceptibly into the other" (p. 7). These concepts were explored below.

1.2.1. General Intelligence: What the Raven's Progressive Matrices measure?

General intelligence or *g* has been widely studied over the last decade. Tests of *g* are useful as they are designed to predict an individual's level of school performance. The SPM, which was the focus of this study, was said to be a good measure of g, but there has been much debate about the nature of *g* and what constructs were really measured by the SPM. According to Spearman's (1927) view, general intelligence is made up of three components. Firstly, *g* involves the perception of regular and irregular geometric shapes. Secondly, *g* involves the logical induction of rules made regarding the perceived units of the test items. Finally, Spearman suggested that these discovered rules are then applied to new phenomena or test items.

Most theorists agreed on the existence of *g*, but described it in various ways. Humphreys (1994) defined *g* as the total intellectual repertoire of behavioural responses that an individual has achieved at any particular point in time. Jensen (1994; cited in Carroll, 1997) defined *g* as "some general property or quality … of the brain" (p. 268). Eysenck (1994) put forward a biological interpretation of *g* by studying the correlations of reaction time and psychological measures. Krechevsky and Gardner (1994; cited in Carroll, 1997), however ignored the existence of *g* in their theory of multiple intelligences, and argued that *g* need not be considered when planning curricula.

According to Carroll (1993) general intelligence is made up of abstract intelligence which can be located within the context of his hierarchical theory of intelligence. Carroll's theory has been used to explain and make predictions about various phenomena and contexts. He developed a 3-stratum hierarchical factor model of cognitive abilities. The highest level of his hierarchical theory is known as stratum III which is general intelligence or *g*. One level lower, stratum II, refers to broad intelligence including fluid intelligence, crystallised intelligence, general memory and learning, broad visual perception, broad auditory perception, broad retrieval ability, broad cognitive speediness and general psychomotor speed. The lowest level is stratum I which refers to narrow abilities such as sequential reasoning, quantitative reasoning, verbal abilities, memory span, visualisation and perceptual speed. Lynn & Irwing (2004) argued that analogical reasoning is made up of verbal-analytic reasoning, which is the same as Carroll (1993) conceptualisation of fluid ability and visuospatial ability. Carrol (1993) defined visuospatial ability as "the ability to rapidly perceive and manipulate visual patterns or to maintain orientation with respect to objects in space" (p. 16).

According to Primi (2001), fluid intelligence (Gf) is strongly associated with the central executive component of working memory that is linked to controlled attention and selective encoding. Fluid intelligence "involves making meaning out of confusion; developing new insights; going beyond the given to perceive that which is not immediately obvious forming (largely non-verbal) constructs which facilitate the handling of complex problems involving many mutual dependent variables" (Raven, Raven & Court, 1998, p. G4).

A considerably quantity of research and speculation has focused on the constructs or cognitive components responsible for performance on the RPM (DeShon, Chan & Weissbein, 1995). Items were designed to steadily increase in difficulty as the test progressed (Meo, Robert & Marucci, 2006). It has been argued that the RPM is the purest measure of *g* in existence. Other scholars suggested that the RPM measures inductive ability, fluid ability, working memory, deductive ability, spatial ability (Rohda & Thompson, 2007), pattern perception or non-verbal intelligence (DeShon et al., 1995). Lynn, Allik & Irwing (2004) found that the RPM is not a pure measure of reasoning ability or general ability, *g*. He suggested that while most items measure reasoning ability, the earlier items measure visualisation, and other items measure analogical reasoning.

Kyllonen & Christal (1990) argued that all these various components of the RPM can be explained by one construct, working memory. According to MedTerms (2008) working memory is short-term memory. It is a system for temporary storage and management of information needed to carry out complex cognitive tasks such as learning, reasoning, and comprehension. Working memory involves the selection, initiation, and termination of information-processing functions such as encoding, storing, and retrieving data. According to Meo et al., (2006) solving the matrices involves abstract reasoning, by deducing the relationships between the elements of each item. They also involve the identification of underlying rules and the use of these rules to determine the answer. Despite these debates it was widely accepted that the RPM is an excellent measure of individual differences in cognitive ability, and is useful in gaining information about a child's level of performance at school (Snow, Kyllonen & Marshalek, 1984; cited in DeShon et al., 1995). According to Raven, Court & Raven (1977) the predictive validity of school performance of the SPM is high, 0.7. However Anastasi (1988) later argued that the correlation between the SPM and academic criteria was somewhat lower than usual intelligence tests.

Furthermore, there has also been much contention in research regarding the dimensionality of performance on the RPM (DeShon et al., 1995). According to Dillon, Pohlman & Lohman (1981; cited in DeShon et al., 1985) performance is related to addition and subtraction of elements as well as the detection by pattern progression. According to Alderton and Larson (1990) and Arthur and Woehr (1993) a single factor model adequately describes the response patterns on the RPM. Hunt (1974; cited in DeShon, 1995) developed a multi-dimensional conceptualisation of performance, including two general problem-solving algorithms. The first he described as visual strategy, where a person uses operations of visual perception, such as continuing lines through black areas and superimposing visual images upon each other. The second, he argued, is the application of logical operations to features contained within parts of the matrix. Carpenter, et al. (1990) argued that performance on the RPM is determined by the ability of the respondent to generate and maintain goal hierarchies in a single, limited capacity store or working memory. Finally Embretson (1993; cited in DeShon, 1995) suggested that tasks presented on the visual format can be processed using visuo-spatial strategies and/ or verbal-analytic strategies depending on the task stimuli and the goals of the problem solver. Matzen, Molen & Dudink (1994) also pointed out that only the correct choices are taken into account, and thus no performance information can be deduced from incorrect answers. They also mentioned that the SPM manual (Raven, Court and Raven, 1998) does not contain an error analysis, which could also be useful in drawing conclusions about the nature of performance on the SPM.

Van der Ven and Ellis (2000) argued that it is crucial to assess whether the SPM is unidimensional, or that responses to all subtests depend on the same underlying trait or ability. They suggested that if the subtests of the SPM deviate from each other, the test may measure a combination of dimensions. If each subtest contains two or more clusters of items, then it may be inappropriate to infer a theoretical interpretation from the subtest scores impacting the construct validity of a test. This means that unidimesionality of the total test depends on the unidimesionality of each subtest.

Much research regarding the design features of the Raven's Progression Matrices has identified aspects that affect the difficulty of the items. The major source of individual differences may be related to working memory capacity with those obtaining higher scores more able to plan, co-ordinate and monitor large numbers of goals and sub-goals (Meo et al., 2006). Primi (2001) reported a more specific understanding of the component processes of inductive reasoning which is a core measure of fluid intelligence or Gf. Certain factors appeared to affect the item difficulty, which are used to interpret cognitive ability (Arendasy and Sommer, 2005).

Primi (2001) identified four design features or "radicals" in the RPM, also known as complexity factors. These were cognitive capacities that a respondent must possess in order to deal with the test demands and problem solving of parts of the matrices. Each complexity factor involves the demand for more essential cognitive capacities that make up Gf. The first two features were known as the number of elements and the number of rules or the necessary transformations of given matrices. These two radicals were associated with the amount of information stored and processed in working memory (Baddeley, 1986, 1990). Numerous studies have confirmed a link between measured of working memory and inductive reasoning (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1990; Kyllonen et al., 1990; Su¨ß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002).

Several other studies have indicated that the respondents were able to reduce the load of working memory by resorting to a response elimination strategy (Bethell-Fox, Lohman & Snow , 1984; Primi, 2001). This occurs when the correlation between measures of inductive reasoning and measures of working memory capacity considerably decrease (Kahl, Beckmann &, Guthke, 2003; cited in Arendasy et al., 2005). Unsworth & Engle (2005) argued that there is a body of research that suggested a correlation between working memory and general intelligence, but the exact nature of this correlation is unknown. With regard to the SPM, this correlation was fairly consistent across the first three quartiles of difficulty, but decreases considerably for the hardest problems of the test.

The third design feature of Gf required for the correct completion of the RPM was the type of rules required in test items. These rules vary in their difficulty and more difficult rules were more demanding on working memory capacity than easier rules. Primi (2001), argued that the processing time of respondents increases with an increase in the number of elements and the number of rules in a test item.

The fourth radical was perceptual organisation, which involves perceptual features of the elements of a figural matrices item as well as perceptual groupings of elements based on the Gestalt principles of proximity, similarity, common region and continuity (Rock & Palmer, 1990). These two aspects of perceptual organisation were distinguished by the effect that they have on the solution process of the matrices items, or based on the extent to which some sort of abstraction is required. Some perceptual manipulations affected the encoding process as the respondent created an internal mental representation of the aspects of the elements. Other perceptual manipulations affected the correspondence finding process by bringing about a conflict between perceptual and conceptual groupings of the individual elements of the matrices items. This in turn increased the likelihood of the formation of irrelevant groups of elements—thereby increasing the item difficulty (Embretson, 2002; Primi, 2001). This result was in line with earlier research conducted by Hornke and Habon (1986; cited in Arendasy et al., 2005). Habon, (1981) and Hornke & Habon (1984; cited in Arendasy et al., 2005) argued that respondents have differing ability levels in perceptually extracting the individual elements when there are excessive overlays in a test item. This challenged the Rasch Model, which assumed that the influence of the perceptual feature overlay on the difficulty of items can be generalised across different sub-groups of the sample.

Gestalt principles also appeared to apply to the test items on the SPM. The Gestalt principle of “proximity” implied that a test item is more difficult when the fused elements are governed by different rules, thus the individual is required to separate the given shape into its individual elements to solve the item. The solution process required further perceptual effort that may be solved in different ways by different subgroups (Arendasy et al., 2005). The Gestalt principle of "similarity" (figure distortion) was a method to determine the amount of abstraction needed to solve figural matrices (Rock & Palmer, 1990; Embretson, 2002). The third Gestalt principle was "common region", which referred to the positioning of the element within a figural matrices item (Arendasy et al., 2005).

Nettlebeck (1998) reported on Jensen's Galtonian notion that performance of elementary cognitive tasks (ECTs) assessed the speed of information processing, which provided some explanation regarding broad general intelligence. Nettlebeck (1998) highlighted Jensen suggestion that there may be a correlation between reaction time (RT) and IQ tests that measure fluid intelligence such as the RPM.

1.2.2. Intelligence: The Genetic vs. Environmental Debate

According to experts individuals differ on *g*, which is stable over a lifetime and is affected by both genetic and environmental factors. The issue of genetic vs. cultural and / or environmental explanations for differences in intellectual ability was laden with controversy and debate (Hernstein & Murray, 1996; Jencks & Philips, 1998; Jensen, 1998; cited in Skuy, Schutte, Fridjhon, O'Carroll, 2001). Many studies maintained that there is evidence for the genetic model and argue intelligence is hereditary, while others have disputed these arguments (Rushton, 2008). Pal, Shyman and Singh (1997) conducted a twin study that concluded a moderate to high heritability of general intelligence, which was consistent with other studies conducted on the subject.

Historically, Black Americans have gained lower test scores on IQ test as compared to White Americans. This could have been due to poorer educational facilities rather than by what was presumed to be a genetic inferiority. There was evidence for both environmental and genetic factors having influence on intelligence but the debate calls for further research on the subject (Rushton, 2008).

Researches such as Carroll (1997) maintained that general ability should not be accepted as a measure of hereditary intelligence, but rather as a measure of process over a lifespan in achieving mental development. He argued that it is wrong to interpret an IQ score as a measure of individual genetic inheritance without considering the environmental factors. Bock and Moore (1986) viewed intelligence as the result of opportunities presented to an individual by their family, school and everyday experience. The environmental argument held that the norms of mental development intrinsic to mental age and the intelligence quotient result from the fact that on average, children are raised in their native language in "typical" families. They then go to school at a young age, and are presented with "typical" curricula that expose them to reading, writing and numeracy, and they later gain exposure to science and social studies. These tasks and others allow them the skills to perform in different tasks at different ages, as seen in intelligence tests. This developmental journey often occurs in developed countries such as The United States and Britain, where there are broad similarities in family structures and educational systems. Correctly answering items on intelligence tests is reliant on learned skills obtained from daily life and school experience (Carroll, 1997). This view was supported by Nettlebeck (1998) as he argued that socio-cultural factors linked to certain groups may inhibit intelligence.

The "normal" development as measured by IQ tests may be different in cultures where family and education systems are structured differently (Cole & Means, 1981). However, intelligence testing asserted that intelligence develops independently from family and school structures. In contract, Ceci (1991) suggested that education significantly influences IQ formation and maintenance. This appeared too simplified as children from the same educational backgrounds, have varying IQ ranges. Carroll (1997) argued that IQ information does not impact a child's ability to learn. He suggested that all individuals can be trained or educated to a certain extent, and individuals vary in their rate of learning and level of mastery they are able to accomplish.

1.3. Psychological Assessment, Reliability, Validity & Statistical Bias

Psychological assessment is useful in providing information to direct people and organisations in understanding others and to make informed decisions about their functioning. Tools were developed to provide a way to assess human behaviour. These tools included tests, scales, measures, instruments, procedures and techniques (Foxcroft et al., 2004). According to Gregory (2007), a test is used to sample behaviour in order to describe it with categories and scores. Psychological assessment is a process-orientated activity whereby assessors gather a wide array of information and then evaluate it to reach conclusions and make decisions (Foxcroft et al., 2004).

Psychological tests can be useful as they can measure levels of performance and they are potentially flexible between samples and population groups. However, there are three areas of concern with regard to psychological tests (Foxcroft et al., 2004).

Firstly, does a test measure the intended aspect of behaviour consistently? This refers to the reliability of a test. Reliability is a measure that refers to the consistency and repeatability of test scores (Gregory, 2007). However, consistency always implies a certain error in measurement. A person's performance on a test item can be affected by chance factors that are present during the assessment, such as fatigue, noise etc, which may negatively affect a person's performance in a test (Foxcroft et al., 2004).

Secondly, does the test measure what it claims to measure? This looks at test validity. Validity refers to what a test measures and how well it does so (Foxcroft et al., 2004). Validity looks at to what extend the inferences made by a test are appropriate, meaningful and useful (Gregory, 2007). Content validity means that test items represent the kinds of material that they are supposed to present. Criterion validity refers to the degree to which the test correlates with one or two outcome criteria. Construct validity refers to the degree to which the test measures the construct it claims to measure (Murphy and Davidshofer, 1998).

Thirdly, does the test measure what it claims to measure across different groups of people? This is a question of bias in a test where there are systemic errors in predicted performance. Bias can exist in terms of test content, norms or even as a result of the testing situation or context (Israel, 2006). There are a number of types of bias that are often examined in relation to the test content. Construct comparability or conceptual equivalence (Retief, 1988), monitors whether the constructs or ideas in the content, have the same meaning across different people or groups (Owen et al., 1996). This form of bias is related to issues of language and translation, such as similarity of metaphorical or abstract notions (Israel, 2006). With administration, the validity of a measure can be compromised by language difficulties if there is a communication problem between the assessment practitioner and the test-takers. Item format refers to the type of question used in a measure. This can be a problem if the test taker is not familiar with the format of a test question. Many measures consider speed to be an indicator of intelligence; however some cultures see quick responses as culturally inappropriate, thus seriously disadvantaging those learners (Foxcroft et al., 2004).

Furthermore, a construct in one language may be untranslatable in another, thus causing the test-taker to misunderstand the content being referred to (Gregory, 2007). Some languages do not have the concepts and expressions required by tests and thus translation is problematic. An additional problem is that some individuals, particularly in the South African context, speak in a combination of languages known as 'township patios', where a pure version of one language is seldom spoken. An individual who communicates in such a way mixes the different languages they have been exposed to in their interactions. A child who communicates using the patios would be disadvantaged if an assessment was formally translated. Gender bias may also be a factor in a test if certain items favour one gender over the other (Foxcroft et al., 2004).

Predictive bias examines whether a test will predict performance on an independent criterion equally between groups. If this type of bias is present in a test the implications are that the test will systematically overestimate or discriminate in relation to the performance of a particular group, causing an error in measurement of an innate difference (Owen et al., 1996). Learners within the South African context may be disadvantaged during psychological assessment administration if they are not familiar with the kinds of questions being asked such as those within the RPM.

By examining the internal consistency of a test, validity and bias can be explored (Murphy & Davidshofer, 1998). Internal consistency estimates are a function of the number of test items and the average intercorrelation among these items (Murphy & Davidshofer, 1998). The internal consistency of a test looks at how well a test covers a range or an ability for which the test is intended, which refers to validity, and whether the items scale in the same way for different populations, which refers to bias (Raven et al., 1998).

There was widespread belief that tests of cognitive ability may be biased if there are systemic differences in test scores due to socioeconomic status, gender and race (Jensen, 1980; Linn, 1982; cited in Murphy et al., 1998). In general, individuals from middle and upper classes scored higher on tests than individuals from lower socio-economic classes. White people receive higher scores than other groups, and males often receive higher scores than females. This is also a question of validity within cognitive test, such as the SPM, as there may be doubt about whether the test measures what it claims to measure, or whether the test makes systematic errors in measurement or prediction (Murphy et al., 1998).

Cultural background can impact the entire process of assessment as culture affects a person's life views, family roles, problem-solving strategies etc. (Sattler, 1988; Gregory, 2007). It is culture and experience that are most likely to lead to the differences between groups (Owen and Chamberlain, 1996). Thus culture or grouping can be a source of bias (Israel, 2006). Item bias or score comparability is when people of equal ability from different backgrounds do not have an equal opportunity for answering an item correctly. Different groups may also answer items in a specific way (Retief, 1998). Cultural bias refers to the differences in the extent to which a test-taker has had the opportunity to know and become familiar with the specific subject matter or specific processes required in order to answer a test item (Eells et al., 1951; cited in Murphy & Davidshofer, 1998).

Under the umbrella of culture is the notion of acculturation, which is a current issue in South Africa. Acculturation is the process by which people are assimilated into a particular culture. This process occurs at different speeds. For example a person who moves away from a rural community into a city would become acculturated to their new surroundings. This implied that such a person will adapt to their new surroundings by adopting some of the cultural traits of the society around them. This has implications in assessment in South Africa, as a measure may be culturally appropriate to a particular group, but not for other people groups. However, as individuals from other cultures move closer to the culture and values of the group on which the test has been normed perhaps the more appropriate the measure will become (Foxcroft el al., 2004). For this reason, this research chose to use language as the independent variable of analysis and not culture as the sample comes from a government school where the children have a mixture of cultures, languages and background. It would be inappropriate to use the variable of culture as the boundaries between the cultures are debatable. Many of the children live affiliated to a number of cultures. They adopt cultural behaviours and practices of a number of cultures depending on their surroundings and various cultural influences. Language is a more definable variable as the learners will be either speak English as their first language or not.

Human characteristics are measured in psychology by focusing on a particular group or sample of a population. These studies are scored and said to be normally distributed for a population (Foxcroft et al., 2001). Norm scores are used as a comparative frame of reference (Gregory, 2007). A standard normal distribution, known as the bell-shaped distribution, has a mean of 0 and a standard deviation of 1 (Foxcroft et al., 2004). Standardised assessment instruments are appropriate to the population from which they were standardised and norm-reference tests have the tendency to be biased for other people groups within the population (Beech & Singleton, 1997). This is important as the SPM may not measure non-verbal intelligence in the same way for different South African language groups.

Over several years, many normative studies were carried out in different parts of England, Ireland and the United States with regards to the RPM. Until 1979, there was a noticeable uniformity in the normative scores. This shifted from 1979 onwards, where the RPM was widely normed in a number of countries and for different age groups. Norms in Germany, Czechoslovakia, China, Taiwan, and New Zealand were well above those obtained in England (Raven, 1989). This demonstrated that performance on the RPM was different for different countries and for different population groups.

If there is evidence that a test is not reliable or consistent, valid, or biased it raises questions about the usefulness of the test. Research in this area is important as it looks at the accuracy and fairness of tests across different groups of people. This issue is ongoing in South Africa due to the effects of segregation and discrimination under Apartheid. Tests that have been standardised according to a western population may not be appropriate to other populations in South Africa due cultural and linguistic differences as well as previous socioeconomic and educational disadvantage. According to Asmal (2001), systems should be developed in South Africa that accommodate and respect diversity. This applies to psychological assessment as scores may favour western individuals, and research is important in assessing whether certain tests are appropriate for different people groups in South Africa.

1.4. Assessment and Education in the South African Context

Psychoeducational assessment aims to determine what individuals have, or have not, learned. Education and assessment have been contentious issues in South Africa. Factors such as the level and quality of education, access to resources and learners' ability to read and write have dramatically impacted psychometric test scores (Israel, 2006). Thus, there has been much controversy in South Africa surrounding psychoeducational assessment as tests have been criticised both internationally and in South Africa (Amod, Skuy, Sonderup & Fridjhon, 2000). Foxcroft, Paterson, le Roux & Herbst (2004), argued that psychological assessment in South Africa needs intensive, large-scale development and adaptation as well as test revision if psychological assessment procedures are to be ethical and culturally sound.

It has been argued that assessment may be culturally bound and thus unfair toward different cultural, ethnic or linguistic groups (Bryans, 1992; Gipps, 1990; Gregory & Kelly, 1992; Joyce, 1988; cited in Beech & Singleton, 1997). Standardized tests, especially intelligence tests, have been criticised in this regard (Cummins, 1984; cited in Beech et al., 1997). It is therefore important for an assessor to understand a child's cultural and linguistic background before assessing the child's needs and making judgments about the child's mental ability and functioning (Beech & Singleton, 1997).

Rushton et al. (2000) also reported that due to the previous Apartheid regime in South Africa, many Black South Africans experience greater unemployment, poorer schools, libraries, and study facilities, than do their White counterparts. Many Black South Africans live in overcrowded homes with little to no water or electricity, and suffer from poor nutrition.

Black South Africans may have less exposure or stimulation to the constructs measured on IQ tests, resulting in poorer performance. IQ tests have been seen as biased toward Black children as IQ tests may measure acculturation into western society rather than intelligence (Sattler & Gwynne, 1982; Williams, 1975, in Vincent, 1991).

However, research has shown that fully bilingual children are not generally disadvantaged when it comes to psychometric testing (Cummins, in Beech et al., 1997). With this in mind, Rushton et al. (2000) argued that instead of labeling standardised testing as “racist” in South Africa, intensive research should be conducted to develope norms for different South African groups, recognizing a variety of talents in people, as well as teaching problem solving techniques and assessing the effectiveness of mediation.

According to Owen (1992), SPM may not be culturally or ethnically “blind”. Owen (1992) found that mean scores between Black and White South Africans was almost 3 SD units and Black and White South African show a general pattern in test answers and performance in their own group, which are dissimilar to that of the Coloured and Indian groups. Owen (1992) argued that these differences pose serious problems for psychologists who wish to establish common tests all for people and therefore all South Africans.

This research investigated these assumptions by examining whether SPM is biased in terms of language and gender in South African high school children. This research also investigated the reliability and validity of the SPM in the South African context. The relevance of the SPM which was normed in Britain, USA and other countries was examined in the multilingual and multicultural population in South Africa.

1.5. Studies concerning the Raven’s Progressive Matrices

This research investigated the relevance of the SPM in the South African context with regard to language and gender. Gender and language are issues of concern in the South African context as psychometric intelligence test are being used throughout the population groups.

This section explored the literature and the debates therein concerning the SPM with regard to language and gender.

1.5.1. Standard Progressive Matrices: Language and Group Differences

Language is a pertinent issue with regard to intelligence testing in South Africa. There are 11 official languages with many people fluent in a number of these and other languages. Perhaps there are differences among the language groups with regard to how individuals approach the constructs of the SPM. If there is no language bias in the SPM, it could be a useful predictor of school performance across the groups.

Over a thousand studies of Raven's Progressive Matrices have been conducted over the last century (Moran, 1986) and many of the studies concluded that the test is culturally and linguistically useful for diverse groups of people (Valencia, 1984). Other studies have argued that test scores on the RPM are influenced by language, motor and sensory abilities and socio-economic factors (MacArthur and Elley, 1963). A study conducted in Egypt (Abdek-Khalek, 1988) reported a high concurrent validity with regard to the SPM as mean scores of Egyptian males were similar to those from British males, although the mean scores of the Egyptian females were lower. Thus, the study concluded that the SPM is a viable tool in the Egyptian context.

According to Raven (1989) historically test scores were different among different ethic groups, favouring individuals of the western culture. He argued that more recently, differences in scores are less obvious as children from different backgrounds are developing the ability to perceive and think clearly. Raven (1989) argued that increasingly, parents and teachers are encouraging children to conceptualise things for themselves. Skuy et al. (2001) also maintained that more recent phenomena that the IQ scores in industrialised societies are rising.

For psychologists in South Africa, the Euro-American psychometric test norms are a concern for testing people from an array of ethnic groups, with regards to test validity. In general, lower means scores were obtained in African samples relative to Euro-American test norms (Grieve & Viljoen, 2000; Owen, 1992; Rushton & Skuy, 2000). Snyderman and Rothman (1988) argued that by examining genetic theory, there should be a reasonably high degree of consistency in the intelligence levels across populations and geographical locations and there might be some genetic basis to the low IQ scores of Black individuals.

One of the earliest studies done in South Africa with regard to the SPM was a study by Notcutt (1949-50). The test was applied to Zulu primary school children. Notcutt found the Zulu scores between 3 and 5 years below the British norms with differences becoming larger among older children. This is similar to Jensen's (1974, 1976) study that hypothesised "cumulative deficits in intelligence test performance in low-SES black children" (Owen, 1992, p. 150). Much later Rushton, Skuy & Fridjhon (2002) conducted a study among African, Indian, and White engineering students in South Africa and found Jensen effects showing pronounced differences in scores among the groups. They argued that test takers should be similar in cultural, educational and socio-economic background for the scores to be compared. Because the SPM was considered biased in this case, it was argued that "true" African mean IQ should be determined. Research should also examine whether African/ non-African differences are on the *g* factor, whether the IQ scores obtained by African learners were predictive of their school performance, and whether intervention techniques would raise their IQs.

Lynn's (1991) review of 11 studies recounted an average IQ of 70, for East, West and Southern Africans, as compared to average IQ of 85 for Black Americans and an average IQ of 100 for Whites. Lynn (1978) also reported consistent results of seven other African studies showing pupils to have IQ equivalents ranging from 75 to 88, with a mean of 82. Lynn (1991) challenged previous genetic theory and stated that IQ tests are biased towards African people, because African people have experienced adverse social and economic living conditions. Furthermore, Crawford-Nutt (1976) administered the SPM to 114 high school pupils living in Soweto, Johannesburg. It was found that, when given clear instructions about the task, Grade 11 and 12 performed similarly to the normed group. This questioned the genetic theory about intelligence and the author concluded that the low scores often obtained by Black South Africans were not necessarily a reflection of their ability, but rather the result of the method of test administration. Rushton & Skuy (2000) maintained that further research was necessary in South Africa to normalise distributions for the African populations on existing tests, and to develop new tests that included social intelligence.

Skuy, Hoffenberg, Visser and Fridjhon (1990) and Skuy and Shmukler (1987) reported that low performance on IQ tests could be improved through mediation resulting in improved performance on the SPM. Rushton and Skuy (2000), suggested that instead of focusing on the impact and results of intelligence testing, talents of pupils should be identified and nurtured. Methods needed in the South Africa context were mediation, as well as the distribution of vitamin and mineral supplements to enhance children's learning potential. They argued that examining these issues would tell researchers a lot about the nature as well as the nurture of intelligence.

Rushton, Skuy and Bons (2004) investigated the construct validity of the RAPM for Black and Non-black South African engineering students and reported that the Raven's Matrices scores were as valid for Black South Africans as they were for non-Black South Africans, thus reflecting the *g* factor of intelligence rather than any cultural way of thinking. Rushton et al. (2004) stated that the RPM and other *g* loaded tests were appropriate for highly educated Black Africans.

It has been argued that performance on the RPM is relatively uninfluenced by culture, and does not require language mediation, specific or previous knowledge. However, Caffarra, Vezzadini, Zonato, Copelli and Venneri (2003), argued that performance in the shorter four-set (A, B, C, D) version of the RPM was significantly affected by age and education level, but not by gender.

More recently, Israel (2006), reported a substantial language bias at item level for university students in South Africa. Israel suggested that the ability to comprehend English had a relatively strong impact on performance on the RAPM as items were not equally difficult across language groups, negating the linguistic fairness of the test as previously suggested. Israel (2006) also suggested that African first-language speakers may be more likely to systematically fail to discriminate important from unimportant information in the test items, indicating a bias at the level of individual items. The implications of Israel's study were that either substantial training is necessary to make the test fairer between different language groups, or alternate norms need to be developed. Israel (2006) argued that the RAPM should be administered with extreme caution as her study contradicts previous notions that the test was culturally fair.

With regards to the SPM specifically, Owen (1992) compared the performance on the SPM of White, Coloured, Indian and Black Grade 9 students in South Africa and found that the measure behaves the same way for different groups, but with large mean differences between the Black and White groups implying that the norms should be different between different groups. He argued that children from different groups may not use the same solution strategies (Foxcroft et al., 2004). Rushton et al. (2000) also found that African university students who had been accepted into university after a battery of tests, scored lower on the SPM than the White university students with equal ability, showing some language bias on the test.

Caffarra et al. (2003) suggested that norms on the RPM should be periodically updated to ensure that normative values are truly reflective of contemporary people's abilities, as there have been cultural and educational changes across generations. Sundberg & Gonzales (1981) caution that the SPM has been seldom validated against relevant criteria in other cultures and therefore should be used with caution. This was particularly important in South Africa as there have been major changes in the education system since 1994. It is therefore important to establish whether the same constructs measure academic performance and non-verbal intelligence equally for different cultural and language groups (Foxcroft et al., 2004).

A study comparing 8- and 12- year olds Black children from rural communities (Venda and Transkei) to urban children (Soweto) found that the urbanised children performed better than the rural children on the SPM (Freeman, 1984). The author concluded that cognitive style needed to deal with the SPM is better developed within an urbanised population than a rural one. This again was evidence that environmental factors affect performance on the SPM. This is different to the current study as the sample used in the current study have had at least five years of education with English as the language of instruction. The environmental factors discussed above then would have been ruled out.

1.5.2. Standard Progressive Matrices: Gender Differences

Much international research and research in South Africa has investigated the gender differences in the SPM. There was a diversity of findings, but more recent literature has shown an interesting developmental trend in SPM scores and general intelligence across different populations and nations.

For over a century there was consensus among researchers that there was no significant difference between males and females on the SPM, and therefore no gender difference in general intelligence (Brody, 1992; Cattell, 1971; Court; 1983; Halpern, 2000; Jackson & Rushton, 2006; Jensen, 1998; Mackintosh, 1996; Spearman, 1923). Colom, Garcia, Juan-Espinosa and Abad (2002) found no sex difference in their study, but their argument was contested by Nyborg (2005) due to the nature of the statistics Colom et al. applied. However, males have reportedly higher average scores on some tests of spatial ability, mathematical reasoning and targeting, while females appear to average higher scores on tests of memory, verbal ability, motor co-ordination and personal space (Halpern, 2000; Kimura, 1999). Males' scores in many of these abilities showed greater variance with scores on both the high and low extremes (Deary, Thorpe, Wilson, Starr & Whalley, 2003; Hedges & Nowell, 1995; cited in Jackson et al., 2006). In more recent years there have been marked differences found between males and females with IQ favouring males, especially in adulthood (Lynn, 1994, 1999; Lynn & Irwing, 2004; Irwing & Lynn, 2005).

Initially, if a marked gender difference was found in a particular ability, this was deemed a gender-bias and such items were then excluded from general test batteries (Kimura, 1999; Voyer, Voyer & Bryden, 1995). Later it was argued that gender differences favouring males on intelligence tests were the result of brain size, as men have comparatively larger brains than women. Many studies (e.g. Ankley, 1992; Jackson & Rushton, 2006; Rushton, 1992) claimed that there was a correlation between brain size and IQ. Murray (2003) pointed out that perhaps males are intellectually superior to women as men have contributed to 98% of the world's historical knowledge.

Gender differences in the RPM have also been widely researched. Colom, Escorial and Rebollo (2004) reported that 87 previous studies showed males out-perform females on the RAPM. Certain test items appeared to favour males on the RAPM due to males having a more superior spatial ability as reported by Abad et al. (2004) on their study of university students in Spain. This suggested a gender bias in the RAPM due to the visual-spatial format of the test. Mackintosh & Bennett (2005) reported significant differences in IQ scores on the RAPM favouring males in American college students. They suggested that this may be due to an unusually high proportion of items that favour males. Colom & Garcia-Lopez (2002) argued that evaluation of sex differences on *g* must not be based on a single test as results would be inconclusive. The results of their study strongly suggested that there were no gender differences on *g*.

With regards to the SPM, Lynn (1994, 1997) reported no significant gender differences on the SPM from 5 to 15 years of age based on a study of 12 datasets from United States, Britain, Norway, Sweden, Indonesia, Northern Ireland, the Netherlands and China. However, Lynn (1994, 1999) reported a male advantage of about 4 IQ points for adults. Lynn (1999) studied a further 20 data sets from United States, Portugal, Japan, Scottland, South Africa, Estonia, Germany, Greece, Ireland, Israel, England, Hawaii and Belgium only to find similar evidence for gender differences in intelligence. These finding were confirmed by other studies completed in Spain (Colom & Garci-Lopez, 2002), Kuwait (Abdel-Khadek & Lynn, 2006) and Mexico (Lynn, Backoff, Contreras-Nino, 2004).

Later Lynn & Irwing (2004) and Lynn, Allik & Irwing (2004) reported a developmental theory regarding intellectual development. According to this theory girls develop and mature both physically and mentally over the age range of 8-14 years. From the age of 15 the growth of girls decelerates while that of boys continues. Lynn, Raine, Venables and Mednick (2005) and Colom and Lynn (2004) found the performance of girls to be higher than boys from as young as 3-years old. Girls then do relatively well until the age of 14 years, with scores on the SPM equal to or better than males of the same age, but with little difference between the ages of 13 and 15 years. Finally, males appeared to out-perform females on the SPM from 15 years into adulthood, where adult men obtained higher mean scores (between 5 and 8 IQ points) on the SPM then adult women, which indicated a higher *g* for men than women (Lynn & Irwing, 2004; Nyborg, 2005).

Females tend to develop earlier than males and thus show earlier advantage on the SPM. Males appear to only develop visualisation and spatial abilities from the age of 12, and their verbal-analytic reasoning and visuospatial ability continue to develop up to the age of 17 (Lynn & Irwing, 2004). From the age of 16 the growth of girls decelerates when compared to boys, thus males have larger average brain size compared to females (Colom & Lynn, 2004). Ankey (1992) and Rushton (1992) agreed that males have greater average brain size than females, in relation to body size by approximately 100 grams with a standard deviation of 128 giving a male advantage of 0.78d. This should therefore give males approximately 4.1 IQ points advantage over females. With this phenomenon of IQ developmentally shifting advantage between the genders and ultimately favouring males, it has been suggested that the male advantage does not emerge until late adolescence when the brain size of males begins to peak. Perhaps girls mature faster than boys giving them an early language advantage, which may mask their cognitive differences (Jackson et al., 2006).

This profile of sex differences in abilities was apparent in the Spanish sample investigated by Colom and Lynn (2004), and resembled the gender differences found in the United States and Britain, which showed a robustness of the developmental theory of intelligence with regard to gender. Jackson and Rushton (2006) found that gender differences favouring males on the SPM were apparent in many different socioeconomic and ethnic groups.

Lynn (2002) investigated the gender differences on the SPM on a very large sample of 15 and 16 year olds in South Africa and found some spatial advantage in males, especially in later adolescence. With a sample of 3979 15-16 year old South Africans, Lynn (2002) supported the developmental theory that males have an advantage over females after the age of 15 years. This finding was confirmed by a study conducted by Flynn (1998). Rushton and Skuy (2000) also reported a small gender difference favouring males on the SPM, but unrelated to *g*.

Language is an important issue in education and assessment in South Africa, as children with the same educational backgrounds are not making similar progress in schools, and the reason could be language related. Test scores on the SPM may be confounded by language differences, contradicting the notion that such non-verbal intelligence tests are not affected by language, making it an invalid base score of IQ, general intelligence and eductive ability. Gender is also an important factor to be investigated as literature regarding the SPM showed males to have an advantage on the test from the age of 15 years and older. A difference in gender favouring males may either indicate a gender bias or a male advantage with regard to IQ. Differences in scores on certain tasks may be the result of the spatial nature of certain test items.

Literature suggests that there was much contention regarding the appropriateness of the RPM for different groups of people and emphasised how further study into the usefulness of the RPM and specifically the SPM in a South African context is necessary. This research focused on high school children in South Africa who have different language backgrounds but are taught in English within the same school, as there is no previous research on this. This research also examined whether trends in gender differences were apparent, as this has not been widely researched in South Africa, especially in the high school age group.

The linguistic fairness and gender differences of the SPM have not yet been adequately researched in high school children in South Africa. This study investigated whether the SPM is reliable and valid in the South African context. Gender and language differences were also investigated on the overall test scores and on specific test items. Few psychological tests have been standardised for the South African context, and thus research and literature in this regard are important.

1.6. Research Questions

The following research questions were investigated:

1. Is there evidence of language bias in the Raven’s Standard Progressive Matrices in high school children in South Africa?
2. Is there evidence of gender bias in the Raven’s Standard Progressive Matrices in high school children in South Africa?
3. Is there an interaction between the language and gender variables on the scores of the Raven’s Standard Progressive Matrices for South African high school children?

CHAPTER TWO: METHODS

2.1. Research Design

This was a quantitative, ex post facto non-experimental, between subjects research design where variables were not manipulated. Once off measurements were administered for the Grade 8 and Grade 9 learners and information on performance on tests were analysed in terms of language and gender. The current study was a non-experimental study as a pre-existing participant variable was used rather than a manipulation of an independent variable (Gravetter & Forzano, 2003). The independent participant variables defining the groups were language, English First Language (EFL), English Additional Language (EAL), and gender, male and female, which were independent variables, and the Raven's Standard Progressive Matrices, which was a dependent variable. The variables were existing variables, regarding the archival data that had already been collected over a two year period. The descriptive data described the relationship between the variables rather than a cause-and-effect relationship between the variables.

2.2. Sample

The subjects for this research were from a government school in Gauteng. The school was previously a model C school and an English medium school. Approximately one third of the learners were EFL learners, while the rest were EAL learners. This research used two years of data (2006 and 2007) from learners from both Grade 8 and Grade 9. The learners had all been exposed to at least 5 years of education in an English medium school. The number of boys and girls was roughly equal between the EFL and the EAL learners. 161 learners were administered the SPM in both Grade 8 and Grade 9.

Table 1: Demographic Details of Sample

	AGE						GENDER		HOME LANG		GENDER * HOME LANG			
	13y	14y	15y	16y	17y	u/k	M	F	English	Other	M / EFL	F / EFL	M /EAL	F / EAL
Gr. 8 N=414	66	266	73	6	1	1	219	194	162	251	101	51	118	143
Gr. 9 N=358	1	107	202	42	3	2	168	189	161	196	89	107	79	82
											MATCHED PAIRED			
											35	22	43	61

2.3. Instruments

A biographical questionnaire was administered to the learners to gather information about the subjects' language and gender (see Appendix 3a). The SPM was also administered (see Appendix 3b). The SPM were first standardised by John Raven on 1407 children in England, in 1938 (Raven, 1941; cited in Raven, 2003). The SPM is a 60-item timed (20 minute) ability test, measuring eductive ability, which is a component of Gf or fluid intelligence. The 60 items are divided into 5 groups of 12 items, with increasing levels of difficulty within each group (Charmorro-Premuzic, Moutafi & Furnham, 2005). Each test item requires the identification of a missing segment to complete a larger pattern. Some of the items are presented in the form of a 2x2 and others a 3x3 matrix, giving the test its name (Gregory, 2007). By completing the pattern the subjects demonstrate the degree of competence in the reasoning being measured. In general, nonverbal assessments, such as the SPM, attempt to remove language barriers in the estimation of a student's intellectual aptitude. This is especially helpful in assessing South African children with limited language ability in English, the primary language used in most schools (Logsdon, 2008).

According to Raven et al. (1998), over 40 studies, which cover a wide range of population groups, have dealt with the reliability of the SPM. With regard to internal consistency, the correlations between the item difficulties established separately in the UK, US, Germany, New Zealand and China standardisations ranged from .98 to 1.00. According to Owen (1992; cited in Raven, 1998) the test had the same psychometric properties among all groups in South Africa. According to Anastasi (1988) the reliability co-efficient of the SPM was high. The retest reliability with an interval of 1 year between tests was between 0.83 and 0.93, with the internal consistency coefficients between 0.80 and 0.90. Therefore Raven et al., (1998) argued that the SPM is extremely robust and measures the same construct in a wide range of cultural, socio-economic, and ethnic groups despite the variation in mean scores between the groups.

In terms of test-retest reliability the original work showed the SPM to have a reliability range between .83 and .93. More recently, many studies have been conducted showing the reliability to be well above .78. This indicated satisfactory retest reliability for the SPM for periods of up to one year. Beyond a year the evidence was limited mainly due to sample size and loss of respondents from initial samples (Raven et al., 1998). Kendall, Verster and Von Mollenderf (1988; cited in Rushton, 2008) confirmed these high test-retest reliabilities but argued that the external validities from the SPM ranged from 0.30 – 0.50 for Black, Coloured and Indian and White South Africans, which implied that performance on the SPM was not necessarily a good indication of school and everyday performance.

Reliable correlations between the SPM and the Binet and Wechsler scales exist for English speaking children, which indicated good criterion-orientated validity. However, correlations between the SPM and performance on achievement test and scholastic aptitude tests were generally lower (Raven et al., 1998). According to te Nijenhuis, van Vianen and van der Flier (2007) the *g* components of tests, such as the SPM, were responsible for the criterion-related validity. With regards to predictive validity, coefficients reported in studies with English and non-English speaking children and adolescences generally ranged up to .70. The reports of few cross-cultural studies indicated a need for more research in different cultural contexts, with different criteria comparison (Raven et al., 1998). A study conducted by Schweitzer, Goldhammer, Rauch and Moosbrugger (2007) reported the SPM to possess good convergent validity but slightly impaired discriminant validity.

The content validity on the SPM was indicated by correlations, which ranged from .20 to. 80, with a good to excellent discrimination power for most test items. It was observed that the SPM meets the requirements for use in cross-cultural contexts and has face validity as it assesses the basic ability to reason in a way that is not obviously biased. With regards to factorial construct validity, investigations with British children revealed high loading on *g* up to .80 and up to .81 in the USA. The research suggests that the SPM is a relatively good measure of general intellectual ability, but is not a pure *g* estimate, especially in cross-cultural contexts (Raven et al., 1998).

2.4. Procedure

The archival data used was collected two years prior to the current study. Permission to collect and use the archival data was applied for and granted from the Gauteng Department of Education. The principal of the school was given a letter describing the nature and purpose of the research. With the principals permission, the parents and the Grade 8 and Grade 9 learners were sent consent and assent forms (see Appendix 3c, Appendix 3d & Appendix 3e). The letters included details regarding the purpose of the study, the learners' involvement, how the information would be used and feedback to parents and learners. The letter stated that participation was voluntary and no negative consequences would result in non-participation. All information remained confidential and no identifying information was included in the report. Raw data was stored in a secure place. There were no foreseeable risks involved for the participants and feedback was given to the school but not the individual participants. Biographical questionnaires were then sent to the learners from whom consent and assent forms were received (see Appendix 3a). Based on questionnaires the subjects for the current research were selected on the basis of whether they had received at least five years of English medium education.

The sample was then administered the Brown Level of the Stanford Diagnostic Reading Test (SDRT) as well as the Raven's Standard Progressive Matrices. The test was administered in a group setting within their classroom by research assistants. The duration of the tests was approximately 60 minutes, after which the tests were collected, scored and collated. The raw scores, indicating the learners' performance on the tests, and a norm-referenced score were recorded for each test subject.

The data collection adhered to the requirements of the University of the Witwatersrand regarding ethical research with human subjects. Subsequent to the collection of the original data, permission to analyse this archival data for the current study was obtained from the Committee of Research on Human Subjects – Humanities and the Graduate School.

CHAPTER THREE: RESULTS

This chapter presents the summary of the data and an assessment of the reliability of the SPM data. A discussion of the suitability of the data for parametric analysis follows, as well as the statistical techniques that address the three research questions. These statistical analyses consisted of 2-way ANOVAs and item analyses, including an item difficulty, run separately for the Grade 8 and Grade 9 data. A matched t-test was analysed for the subjects who received the SPM in both Grade 8 and Grade 9. These statistical tests were used to draw conclusions about the validity and reliability of the SPM by examining if there is any evidence of test bias on the basis of language and gender.

3.1. Descriptive Statistics

Table 2: Descriptive Statistics for the SPM – Grade 8 and Grade 9

	Gender	Home Lang	Mean SPM Score	Std Dev	Median	Mode	Minimum	Maximum
Gr. 8 Total SPM* Score**	**Overall Data**		42.32	6.49	43	41	12	56
	Male	EFL*	43.62	6.04	44	45	23	56
		EAL**	43.11	6.01	43	42	13	55
	Female	EFL	43.25	5.73	44	44	28	54
		EAL	41.23	7.25	43	41	12	52
Gr. 9 Total SPM	**Overall Data**		44.54	6.07	45	44	10	58
	Male	EFL	45.22	5.18	44	44	32	58
		EAL	44.40	6.92	45	43	10	58
	Female	EFL	44.66	6.53	46	43	17	55
		EAL	44.05	5.57	44	44	31	53

* EFL = English First Language

* EAL = English Additional Language

***SPM = Standard Progressive Matrices

According to Table 2, the overall means of the Grade 8 data were consistent across the groups, with a slightly lower mean for the Female EAL group ($\mu = 41.23$). The overall standard deviation was relatively small ($\sigma = 6.49$) showing an accurate measure of each subjects standing on *g* (Murphy & Davidshofer, 1998). The standard deviations for the four groups were consistent with the overall standard deviation. The minimum scores for both

EAL groups (Male = 13; Female = 12) were considerably lower than the minimum scores of the EFL groups (Male = 23; Female = 12). The maximum scores were consistent across the four groups with slightly higher scores for the male subjects (EFL = 56; EAL = 55) than the female subjects (EFL = 54; EAL = 52) (see Appendix 4a).

Table 2 shows the mean scores for the Grade 9 data were relatively consistent across the groups, with a slightly higher mean for the Male EFL group ($\mu = 45.22$). The overall standard deviation was relatively small ($\sigma = 6.07$) showing an accurate measure of each subjects' standing on *g* (Murphy et al., 1998). The standard deviations for the four groups were consistent with the overall standard deviation. The minimum score for the Male groups differed considerably (EAL = 10; EFL = 32). Unlike the Grade 8 sample, the minimum score of the Female EFL subjects was 17, which was considerably lower than the Female EAL minimum score of 31 (see Appendix 4b).

3.2. Reliability of the Instrument

Internal consistency reliability estimates, as indicated by Cronbach Coefficient Alpha (Howell, 1999), were calculated for both the Grade 8 and Grade 9 data sets.

Table 3: Reliability Estimates for the SPM

	Cronbach Coefficient Alpha	
	Raw Variables	**Standardised Variables**
SPM Grade 8 data	0.78	0.85
SPM Grade 9 data	0.77	0.84

Table 3 indicates that within the study the SPM displayed an equally high level of internal consistency for the Grade 8 and Grade 9 data. ($\alpha = 0.85$ and $\alpha = 0.84$ respectively). The SPM provided an acceptably consistent measurement throughout the data (see Appendix 4c).

3.3. Normality of the Data

To make use of parametric techniques for statistical analysis, it is necessary to establish that the data was normally distributed (Howell, 1999). This was done with the use of histograms, measures of central tendency and Kolmogorov-Smirnov Tests of Normality. All these measures confirmed that the data is normally distributed for both the Grade 8 and Grade 9 data and thus allowed for parametric analyses (see Appendix 4d).

3.4. Norm Scores

The mean scores for each group were compared to the 1986 norms for children and young people in Australia in the context of 1979 British Data. This was done to establish the level of performance of the subjects compared to the standardised population (see Appendix 4e) for the majority age group for each data set. The majority age of the Grade 8 sample was 14 (see Table 1) and the majority age of the Grade 9 sample was 15. Because the SPM was administered at the end of the school year, the 14 years 3 months – 14 years 8 months age was used for the Grade 8 means and the 15 years 3 months – 15 years 8 months age was used for the Grade 9 means on the norms table.

The overall mean for the Grade 8 group was 42.32, which showed the average of the entire sample to be near the 25^{th} percentile compared to the Australian and British norms. The Male EFL ($\mu = 43.62$), Male EAL ($\mu = 43.11$) and Female EFL ($\mu = 43.25$) groups lay between the 25^{th} and 50^{th} percentiles of the Australian and British norms, showing the Australian and British mean scores to be slightly higher than the overall mean scores for the Grade 8 sample from the current study. The mean scores for the Female EAL group lay between the 10^{th} and 25^{th} percentiles, which was considerably lower than the Australian and British norm scores.

The Grade 9 means were slightly higher than the Grade 8 means. The overall Grade 9 mean was 44.54, which showed the average of the entire sample to be between the 25^{th} and 50^{th} percentiles as compared to the Australian and British norms. The mean score of the Male EFL group ($\mu = 45.22$) was slightly higher than the overall mean score, but remained within the 25^{th} and 50^{th} percentile rank according to the Australian and British norms. The Males EAL ($\mu = 44.404$), Female EFL ($\mu = 44.66$) and Females EAL ($\mu = 44.05$) groups were consistent with the overall mean score and also were between the 25^{th} and 50^{th} percentile rank according to the Australian and British norms. This showed that the mean scores of the current sample was lower than the mean scores of the Australian and British sample, raising concern about whether the norms established in Britain and Australia are appropriate for this Grade 8 and Grade 9 sample from South Africa.

3.5. Hypotheses

The following hypotheses were addressed with the use of the statistical analysis that follows:

- There will be a difference in non-verbal intelligence as measured by the Raven's Standard Progressive Matrices between EFL and EAL South African high school children.
- There will be a difference in non-verbal intelligence between Male and Female South African high school children.
- There will be an interaction between the language and gender variables on the scores of the Raven's Standard Progressive Matrices for South African high school children.

3.6. Overall Difference: The 2-way ANOVAs

In order to test the above hypotheses on an overall level, this section examined the differences between the independent variables, gender and language, through the use of the two-way analysis of variance (2-way ANOVA) tests that were run separately on the Grade 8 and Grade 9. The differences on the total Raven's scores were examined first for each data set, followed by a closer examination of the SPM subtests. The 2 - way ANOVAs were run to investigate whether there was a difference between and among sample means. This method looked at whether any number of means differ. The 2-way ANOVA allowed for two or more independent variables to be simultaneously examined (interacting effects), as well as individually examined (Howell, 1999). This was useful in looking at the differences in means on the overall scores between EFL learners compared to EAL learners, as well as the differences between the genders. The 2-way ANOVA also examined the interaction between the two variables, language and gender.

Table 4: Grade 8 2-way ANOVAs

Source	DF	Total SPM F	Total A F	Total B F	Total C F	Total D F	Total E F
Gender	1	0.84	2.66	0.01	9.14*	6.28**	2.19
Language	1	6.71*	5.87**	3.10	0.53	2.02	9.26*
Gender * Language	1	0.14	0.78	0.07	0.44	0.87	0.92

* $p < 0.01$

** $p < 0.05$

Table 4 indicates that there was a significant difference on the basis of home language for the Grade 8 sample on the total SPM score which was consistent with the total of subtests E. A gender difference was noted on subtests C and D however, no interaction between the independent variables was noted.

Table 5: Grade 9 2-way ANOVAs

Source	DF	Total SPM F	Total A F	Total B F	Total C F	Total D F	Total E F
Gender	1	0.50	13.12*	0.00	0.66	2.37	0.32
Language	1	1.20	0.10	0.60	0.12	0.00	3.48
Gender * Language	1	0.02	0.42	1.17	0.45	0.06	0.03

* $p < 0.01$

** $p < 0.05$

Table 5 indicated that the only significant difference for the Grade 9 data was on the basis of gender for subtest A. There were no other significant differences between the groups with regard to gender and home language, nor was there a significant interaction between the two independent variables.

3.7. Post Hoc Test

The post hoc analyses followed the 2-way ANOVAs. The Tukey-Kramer adjustment for multiple comparisons (Rosenthal & Rosnow, 1991) was administered for the total SPM scores as well as all the subtests. This analysis examined where the differences in performance occurred. For the Grade 8 data the post hoc analysis showed a significant difference between the Male EFL learners and the Female EAL learners on the total SPM score, subtest A, subtest C and subtest E. There was also a significant difference between the Male EAL learners and the Female EAL on subtest C. For the Grade 9 data the post hoc analysis showed a significant difference on subtest A only, between the Male EFL learners and the Female EFL learners, and between the Male EFL learners and the Female EAL learners (see Appendix 4f).

3.8. Item Analysis

To investigate the above hypotheses on the item level an item analysis, including an item difficulty, was administered. A Chi-Square analysis was run for each item comparing the performance between the two language groups and between the two gender groups. The Fisher Exact Probability Test was also utilised. This test was useful for 2 x 2 tables of independent observations and shows extreme significance making it a more valid test of significance (Rosenthal & Rosnow, 1991).

An item analysis allowed a close examination of the individual tests, which is essential for understanding why a test shows specific levels of reliability and validity. The reliability coefficients revealed information about the effects of measurement error on test scores, validity coefficients provided information about the accuracy of the predictions that were based on test scores. An item analysis can show why a test is reliable or unreliable and may help in understanding why test scores can be used to measure some criteria but not others. Item analysis may also suggest ways of improving the measurement characteristics of a test. An item analysis elicits three kinds of information: information about distractors, information about the item difficulty and information about item discrimination power (Murphy et al., 1998). A detailed item analysis was conducted to investigate the items that proved more challenging for EAL learners as compared to EFL learners, and Male learners compared to Female learners.

Table 6: Grade 8 Item Analysis – Language

Item	Proportion correct: EFL	Proportion correct: EAL
B3	1.00	96.55**
B6	0.91	80.46*
B7	0.78	68.20**
C8	0.59	45.21*
D4	0.94	87.36**
D10	0.76	66.28**
E2	0.68	0.56**
E4	0.48	0.37**
E5	0.50	0.38**
E6	0.48	0.38**
E11	0.07	0.03**

* $p < 0.01$

** $p < 0.05$

Table 7: Grade 9 Item Analysis – Language

Item	Proportion correct: EFL	Proportion correct: EAL
A4	1.0	0.96*
E5	0.68	0.47*
E6	0.57	0.44**
E10	0.17	0.09**

Table 6 and Table 7 indicate the items that were significantly different for the Grade 8 and Grade 9 sample on the basis of home language. This was done to investigate the hypothesis regarding the difference in non-verbal intelligence as measured by the Raven's Standard Progressive Matrices between EFL and EAL South African high school children on the item level. For both the Grade 8 and Grade 9 samples all the items favoured the EFL group. These items may not be appropriate for EAL learners in the South African context.

Table 8: Grade 8 Item Analysis – Gender

Item	Proportion correct: Male	Proportion correct: Female
A11	0.84	0.75**
C2	0.97	0.92*
C6	0.78	0.69**
C8	0.58	0.42*
C11	0.43	0.32**
D11	0.19	0.31*

* $p < 0.01$

** $p < 0.05$

Table 9: Grade 9 Item Analysis - Gender

Item	Proportion correct: Male	Proportion correct: Female
A10	0.97	0.89*
A12	0.67	0.52*
C8	0.57	0.46
D3	0.87	0.95

Table 8 and Table 9 show the items that were significantly different for the Grade 8 and Grade 9 sample on the basis of gender. This was done to investigate the hypothesis regarding the difference in non-verbal intelligence between male and female South African high school children. For the Grade 8 sample, all the items favoured the Male group, except item D11 which favoured the Female group (p=0.31). For the Grade 9 sample, three items favoured the Male group, but item D3 favoured the Female group (p=0.95).

3.8.1 Item Difficulty

In order to assess score comparability, analysis of item difficulty on the basis of gender and language were carried out separately for the Grade 8 and Grade 9 data. The proportion of correct answers or item difficulty index (*p*) on the 60 SPM questions was calculated separately for the four groups (Male EFL; Male EAL; Female EFL; Female EAL) on each item, noting the higher the item difficulty the easier the item. An item difficulty with a *p* value of 0.5 was optimum, as *p* values clustered around 0.5 showed variability. Items that were considered difficult for a group showed a proportion of less than .40, as scores should contain a spread of items with levels of item difficulty ranging from .40 to .70. (Howell, 1997).

The following tables demonstrated the item difficulty indices of selected items on the SPM subtests for the Grade 8 and Grade 9 samples. The ceiling and floor effects for each subtest were examined, as were the items where one or more groups showed a proportion of less than 0.40.

Table 10 highlights the items that showed ceiling effects as the portions were close to 1.0 across the groups (Items A1-A6 and B1). Floor effects were noted for items C12, E10, E11 and E12 as the proportions of correct answers were close to 0 for these items. An examination of the difficult items (p<0.40 for at least one group) showed item C11 to be more difficult for Female learners. Items C11 and E8 proved to be easiest for the Male EAL group ($pC11$=0.45; $pE8$=0.26). Item D11 was more difficult for the Male learners ($pD11$=0.18). Item E4 was notably easier for the Male EFL group ($pE4$=0.54). Finally the item difficulty indices showed items E5, E6 and E7 to be more difficult for the EAL language learners compared to the EFL learners (see Appendix 4g).

Table 10: Grade 8: Item Difficulty Index

Item Difficulty Index (p)	Male EFL	Male EAL	Female EFL	Female EAL
***p*A1**	1.0	1.0	1.0	1.0
***p*A2**	1.0	0.98	1.0	1.0
***p*A3**	1.0	0.98	1.0	0.99
***p*A4**	1.0	0.98	1.0	0.98
***p*A5**	0.99	0.99	0.98	1.0
***p*A6**	1.0	0.99	1.0	0.98
***p*B1**	1.0	0.98	0.98	0.99
***p*C10**	0.43	0.41	0.29	0.34
***p*C11**	0.39	0.45	0.27	0.33
***p*C12**	0.06	0.03	0.01	0.02
***p*D11**	0.18	0.18	0.29	0.32
***p*D12**	0.11	0.14	0.15	0.14
***p*E4**	0.54	0.37	0.35	0.37
***p*E5**	0.51	0.36	0.49	0.39
***p*E6**	0.49	0.36	0.50	0.38
***p*E7**	0.32	0.26	0.35	0.27
***p*E8**	0.26	0.15	0.19	0.16
***p*E9**	0.11	0.16	0.17	0.10
***p*E10**	0.09	0.04	0.01	0.06
***p*E11**	0.08	0.03	0.03	0.02
***p*E12**	0.10	0.04	0.05	0.05

Table 11: Grade 9: Item Difficulty Index

Item Difficulty Index (p)	Male EFL	Male EAL	Female EFL	Female EAL
A1	1.0	1.0	1.0	0.99
A3	1.0	0.98	0.98	0.98
A6	0.98	0.98	0.98	0.99
B1	1.0	0.98	1.0	0.99
B2	1.0	0.96	1.0	0.98
C10	0.41	0.47	0.34	0.40
C12	0.13	0.07	0.14	0.10
D11	0.17	0.24	0.32	0.23
D12	0.20	0.10	0.15	0.17
E7	0.43	0.35	0.41	0.32
E8	0.27	0.32	0.15	0.26
E9	0.16	0.21	0.18	0.15
E10	0.16	0.12	0.17	0.06
E11	0.06	0.06	0.08	0.01
E12	0.07	0.05	0.04	0.05

Table 11 highlighted items A1, A3, A6, B1 and B2 as showing ceiling effects. Floor effects were noted for items C12, D12, E9, E10, E11 and E12.

An examination of the difficult items (*p*<0.40 for at least one group) showed item C10 to be the most difficult for the Female EFL group (*pC10*=0.34). Item D11 was the most difficult for the Male EFL group (*pD11*=0.17). Item E7 was easier for the EFL groups (*p*>0.40) compared to the EAL groups (*p*<0.40). Finally, item E8 was the most difficult for the Female EFL groups (see Appendix 4h).

The item difficulty indices for both Grade 8 and 9 indicated that the items became increasingly more difficult between and within the subtests as suggested by Raven et al. (1998) and Charmorro-Premuzic et al. (2005).

3.9. Matched Paired t-test

A matched pair t-test was administered to the portion of the sample that received the SPM in both Grade 8 and Grade 9. This group of 161 learners were administered the SPM in 2006 and in 2007. A matched paired t-test or repeated measures is useful when the same subjects have completed a test on two occasions (Howell, 1997). The match t-test was useful to this study to determine whether these subjects would have impacted the scores of the entire sample. The repeated samples make up 39% of the Grade 8 sample and 45% of the Grade 9 sample. The following table demonstrated any improvement or decline in performance by subtracting the Grade 9 scores from the Grade 8 scores. Thus a negative t value indicates an improvement.

Table 12: Matched Paired t-test

Variable	DF	Total SPM t value	Total A t value	Total B t value	Total C t value	Total D t value	Total E t value
All groups	160	-5.06*					
Male EFL	34	-1.89	-0.62	-2.19	-0.63	0.42	-1.41
Male EAL	42	-2.64**	0.60	-1.37	-1.92	-1.69	-2.75**
Female EFL	21	-0.49	0.96	0.00	-0.57	0.65	-1.78
Female EAL	60	-4.18*	-2.69*	-2.38**	-3.19*	-2.37*	-2.70**

* $p < 0.01$

** $p < 0.05$

Table 12 showed a significant improvement overall on the SPM score for the learners who received the SPM in both Grade 8 and Grade 9. Significant improvement was noted overall for both EAL groups. Subtests A, B, C and D showed a significant improvement for the Female EAL only. Item E showed a significant improvement for both EAL groups (see Appendix 4i).

CHAPTER FOUR: DISCUSSION

Over the last century theory and research have been developed regarding psychological assessment tests, to ensure such tests are reliable and valid (Foxcroft & Roodt, 2004). One such test is the Raven's Progressive Matrices (RPM), a psychometric test of non-verbal intelligence, which was originally standardised on a British population. The RPM measures one component of Spearman's 'g', the educative, non-verbal ability (Raven, 1994). Internationally, the RPM has been used widely and was regarded as a good base to assess observation, clear thinking and mental capacity (Raven, 1965). This research focused on one of the three Raven's tests, the Raven's Standard Progressive Matrices (SPM) in the context of South Africa. This test was designed for the general population from ages 6 to 80 (Raven, 2003; Arendasy & Sommer, 2005).

The SPM has been documented as culturally fair on the basis of language, which makes it potentially appropriate for South Africa's population, which is culturally, linguistically and racially diverse (Owen, 1992). Historically practices and political policies in South Africa were discriminatory, which negatively affected the development and education of non-white South Africans. Today these policies no longer exist and an ethos of fairness for all has been cultivated. The SPM is still widely used in South Africa and it is important for research in South Africa to ascertain whether commonly used psychological tests, not standardised on a South African population, such as the SPM, are appropriate across the diverse people groups in the country. This research examined whether the SPM was biased with regard to gender and language in the South African context. Few psychological tests have been standardised for the South African context, and thus research and literature in this regard are important.

With the political changes in South Africa since 1994, English has become the medium of instruction in many schools, creating diverse classrooms where children differ with regard to background and language. A child for whom English is not their first language may experience educational and language difficulties, even when they have received primary education in English for a number of years. These children may not fulfill their potential at school because of their language difficulties. In the same way IQ tests such as the SPM may be biased on the basis of language and unfair for children for whom English is not their first language. Such children may not be familiar with the constructs of the test, which might negatively affect their performance in the test. Thus, language is an important variable to consider when investigating whether the test is biased in the South Africa context.

The current study examined how well the SPM measures non-verbal intelligence in EFL and EAL learners, but the abilities of both groups may be on a continuum as different individuals would have been exposed to English in different ways. This linked to the notion that children from different cultures may be generally improving with time on the RPM as they become more familiar with the constructs of the test. Learners, who are being educated in the same classroom, i.e. receiving the same level of education, do not necessarily have the same abilities due to differences in environment, culture and language.

Gender has also been widely researched in terms of performance on the SPM both internationally and in South Africa. Many studies have suggested that there are no differences, or very small differences, in scores between boys and girls (Lynn & Irwing, 2004, Colom & Abad, 2005). Other studies have shown that there are no gender differences on the SPM before 15 years of age, but a gender difference favouring males appears from the age of 15 onward (Abdel-Khadek et al., 2006; Colom & Garcia-Lopez, 2002; Colom et al., 2004; Lynn, 1994, 1997, 1999; Lynn, Backoff & Contreas-Nino, 2004; Lynn, Allik & Irwing, 2004; Lynn, Raine & Venable, 2005). This research investigated whether the gender bias, previously reported and disputed in research, applied to South African high school children.

The current study was a parallel study to the one conducted by Israel (2006). She examined systematic differences in performance on the RAPM, on the basis of home language and gender in one hundred first-year university students. Israel (2006), reported a substantial language bias at item level for university students. She suggested that strong performance on the RAPM was correlated to good English comprehension, making the test items unequally difficult across language groups. Israel's findings raised doubt about the linguistic and cultural fairness of the RAPM as a non-verbal test of intelligence.

The current study explored whether there is any evidence of bias in the SPM on the basis of home language and gender. Bias was explored by an investigation of differences in performance between learners for whom English is their first language (EFL) and children for whom English is an additional language (EAL) in the South African context. Bias was also investigated with regard to differences in performance on the basis of gender. The research made use of archival data that was collected over a two year period. The research questions investigated were: Is there evidence of language or gender bias, or an interaction between these two variables, in the Raven's Standard Progressive Matrices in high school children in South Africa?

2-ways ANOVAs were run to answer the research questions on an overall level. These were run on the total SPM scores and on the total of each subtest (subtests A, B, C, D, E). Item Analyses were run to investigate language and gender differences on the item level. These two analyses were run separately for the Grade 8 data and the Grade 9 data. A matched paired t-test was also run for the portion of the sample that had been administered the test in both Grade 8 and Grade 9 and therefore were included in both those samples. This investigated whether these learners had improved on the SPM between Grade 8 and Grade 9, and how this repeated measure might have influenced the overall data.

The subjects for this research were a convenience sample of 413 Grade 8 and 357 Grade 9 learners from a government school in Gauteng. The school was previously a model C school and is an English medium school. Approximately one third of the learners are EFL learners where the rest are EAL learners. This research used two years of data (2006 and 2007) from learners from both Grade 8 and Grade 9, and 161 of the learners were administered the SPM in both Grades. The learners had been exposed to at least 5 years of learning in an English medium school. The number of boys and girls were roughly equal between both the EFL and the EAL learners.

The main results of this research showed that there was evidence of language bias in the Raven's Standard Progressive Matrices in high school children in South Africa. This was apparent for the Grade 8 data as the EFL learners outperformed the EAL learners on the test. Overall, there was no evidence of gender bias for either sample, and no interaction was noted between the language and gender variables.

The internal consistency of the SPM was measured using a Cronbach Coefficient Alpha . Both the Grade 8 and Grade 9 samples showed an equally high internal consistency. This is consistent to the high reliability co-efficient reported by Raven et al. (1998) and Anastasi (1988). Thus the SPM was considered a reliable test of *g* (Gregory, 2007; Foxcroft et al., 2004).

Raven et al. (1998) argued that the RPM measures eductive ability and is a good measure of that construct, indicating high construct validity on the test. Evidence drawn from Addendum (1995; cited in Raven et al., 1998) showed that all items measure the same construct, as there is a relationship between the percentage who solved the problem and the total scores. However, it is argued that such evidence of construct validity cannot be acquired from factor-analytic studies of internal consistency.

The overall means for the Grade 8 and Grade 9 data were compared to the norm scores of the Australian and British norms. The mean scores for both samples were below the mean scores of the Australian and British norms. This may indicate that the comparative frame of reference (Gregory, 2007) for the British and Australian populations may not be appropriate for the South African population. This finding would suggest that the SPM should be used with extreme circumspection in the South African context. To use a child's score on the test as an isolated indication of the child's non-verbal ability would be an error in judgement as the Australian and British percentile ranks do not appear to be suitable for South African children in general. This difference is consistent with previous studies that showed the performance on the SPM was different for different countries as compared to the British means (Lynn, 1997). The difference in means scores could be explained as cultural bias as argued by Bryans (1992), Gipps (1990), Gregory and Kelly (1992), Joyce (1998) and Cummins (1984; cited in Beech et al., 1997). However, Cummins (1984; cited in Beech et al., 1997) argued that fully bilingual children, and children for whom English is their first language, are generally not disadvantaged in terms of psychometric testing. Therefore both the EFL learners and the EAL learners from this research should not be disadvantaged as most the subjects have had at least five years of schooling with English as the medium of instruction.

Perhaps the lower mean scores were due to norm scores being inappropriate for the South African population (Rushton et al., 2000). In order for the SPM to be more valid in South Africa, norm scores should be developed for the South African population. However, as noted in the literature, South Africa is comprised of many diverse cultural and ethnic groups (Mayekiso et al., 2007) making the development of norm scores a complex task.

From the analysis of the 2-way ANOVAs, the Grade 8 data showed a strong significant difference on the basis of home language, which was evident in subtests A and E only, the easiest and most difficult subtests. Perhaps the language difference was due to the EAL learners being unfamiliar with the structure of the test or test instructions, thus the significant difference noted in subtest A. This raises concern about bias issues of comparability and conceptual equivalence (Retief, 1988), and may indicate predictive bias (Owen et al., 1996) in the test for the EAL Grade 8 learners. Perhaps the initial instructions did not have the same meaning for the EAL learners. The EAL may not have been familiar with the item format (Foxcroft et al., 2004) and thus not familiar with the format of the test by practically answering subtest A.

The significant difference noted on the most difficult subtest, subtest E, was consistent with the results of the item analysis. The items showed a significant language difference for the Grade 8 learners on the more difficult items such as items B7, C8 and D10. The majority of the significant items were from subtest E, including items E2, E4, E5, E6 and E11. Only four items showed a significant difference on the basis of language for the Grade 9 sample, these were items A4, E5, E6 and E10. This may indicate a difference in familiarity with the construct rather than unfamiliarity with the item format, raising uncertainty regarding the construct validity in the SPM (Murphy et al., 1998). It is likely that many of the EAL learners come from disadvantaged backgrounds and / or their parents were disadvantaged, thus making them less familiar with the range of abilities required for the more difficult questions (Raven et al., 1998). This indicated bias in the SPM for Grade 8 learners in the South African context, which is consistent with historical data that reported a difference in scores among different ethnic groups, favouring those from the western culture (Raven, 1989; Sattler et al., 1982, Williams, 1975, cited in Vincent, 1991). Perhaps this difference will diminish as education improves in South Africa.

Interestingly, the Grade 9 data did not indicate a significant language difference on the 2-way ANOVA. These learners have spent more time in high school and perhaps have had more exposure to the constructs measured in the SPM. Perhaps the learners in general were more able to complete the test in the same way across the language groups with more experience and education. This is consistent with Raven's (1989) finding that more recently differences in scores on the SPM were becoming less obvious in children from different backgrounds. The sample used in this study have all had at least five years of learning with English as the language of instruction and thus the differences in ability between EFL learners and EAL learners appears to be diminishing with time.

However, the language difference for the Grade 8 data may cast doubt on the construct validity of the SPM. If the SPM was a true measure of non-verbal intelligence or general intelligence *g*, as suggested by Anastasi (1988), Jensen (1972), DeShon et al. (1995), no difference should be noted on the basis of language, as the learners ability should vary equally across the language groups. Thus, doubt is raised about the appropriateness of the SPM for certain non-English children below a certain age group in the South African context. It should also be noted that the EAL learners account for two-thirds of the Grade 8 sample, while there was almost an equal number learners in each language group for the Grade 9 sample. The large portion of EAL in the Grade 8 data may have had an effect on the statistical analyses.

The matched paired t-test also showed the most improvement on the SPM for both the Male and Female EAL group on the overall score. This improvement was consistent with the studies of Skuy et al. (1987) and Skuy et al. (2002) who reported a significant increase in scores for the non-white groups, and by implication non-English groups, after the administration of Feuerstein's MLE. However, the learners from the current study were not given specific mediation or training in the constructs measured by the SPM. Perhaps the improvement noted in the current study may have been due the practice effect. These learners may also have gained experience in the constructs measured by the SPM through their interactions both at school and at home, thus accounting for the improvement.

However, less than half of the Grade 9 sample had received the SPM in both Grade 8 and Grade 9, thus the practice effect is not sufficient to explain non-significant result for the entire Grade 9 sample. Perhaps the lack of a significant difference on the basis of language was due to the EAL learners being more familiar with the constructs in the test. This may indicate that the SPM is fair on the basis of language in older children or children with certain experience in the English language. This is inconsistent with Jensen (1993) and Vincent, (1991) findings that the SPM is culturally fair as a language difference was not noted for the Grade 9 sample. However, if the SPM were to be deemed culturally or linguistically fair no difference in language for any age group should be apparent. Experience in an English environment should not affect test scores on a test measuring non-verbal ability.

In addition, the mean score for the Female EAL group from the Grade 8 sample was 41.23 and the mean score for the Female EAL group was 44.05 from the Grade 9 sample. Although these means were the lowest between the groups for each sample, they increase the most between Grade 8 and Grade 9. Although overall the Grade 9 sample also showed higher mean scores for all the groups compared to Grade 8 sample, the greatest improvement was for the Female EAL learners. The matched paired t-test also showed the most significant improvement for the Female EAL group, as this was the only group that showed significant improvement overall and on all five subtests. This improvement was also confirmed by the post hoc test as three subtests as well as the total SPM score showed a significant difference between the Male EFL group and the Female EAL group for the Grade 8 sample, while only one subtest showed this difference for the Grade 9 sample.

This finding may imply that the analytical abilities required to correctly answer the items on the SPM, such as general intelligence (Anastasi, 1988; DeShon et al., 1995), fluid intelligence (Raven et a., 1998), working memory (Carpenter et al., 1990), spatial ability (Ackerman et al., 1993; cited in DeShon et al., 1995; Rohde et al., 2007) etc. can be developed over time and are therefore not static. This could indicate that using once off measures of the SPM in the South African context may not provide adequate information about a child's general ability, especially with children who are from previously disadvantaged backgrounds, and for whom English is not their first language. Perhaps using repeated measures of the SPM could provide useful information about learners' non-verbal intelligence.

Overall, the 2-way ANOVAs revealed no gender differences for the Grade 8 or Grade 9 learners. However, subtests C and D showed a significant gender difference for the Grade 8 learners favouring males, and subtest A showed a significant gender difference for the Grade 9 learners favouring males. Results of the item analysis showed five items that significantly favoured the Male learners: A11, C2, C6, C8 and C11, and item D11 favoured the Female learners for the Grade 8 sample. Only four items showed a significant gender difference, three favoured the Male learners, A10, A12 and C8, and one favoured the Female learners, D3 for the Grade 9 sample. This does not indicate a significant difference on the basis of gender, and thus this research concluded that overall the SPM is not biased on the basis of gender for high school learners in South Africa. These differences do not indicate gender bias on the SPM, but slight advantages were found for each gender on certain test items.

This is consistent with much of the international research (Spearman, 1923; Cattell, 1971; Court, 1983; Brody, 1992; Mackintosh, 1996; Jensen, 1998; Halpern, 2000; Jackson & Rushton, 2006), but inconsistent with the findings of Colom, Escorial & Rebollo (2004) who argued that males out-perform females on the RPM, but their research was conducted on older individuals. Lynn (1994, 1997, 1999), Lynn, Backoff and Contreras-Nino (2004), Colom and Garcia-Lopez (2002), and Abdel-Khadek et al. (2006) suggested a developmental theory with regard to *g*. They argued that there is no significant gender difference with regard to *g* for boys and girls between the ages of 5 to 15, but a gender difference in favour of males can be increasingly noted from age 15 years onwards. This research is consistent with this phenomenon as the majority of the learners in both the Grade 8 and Grade 9 sample were 15 years or younger and thus no significant gender difference was noted. The findings of the current study were difficult to compare to the previous studies conducted in South Africa as the samples from previous studies were somewhat older than the sample of this study (Flynn, 1998; Israel, 2006; Lynn, 2002; Rushton & Skuy, 2000). Perhaps a gender difference would be more obvious for the sample had they been tested at a higher grade.

These findings may also be interpreted as evidence for the environmental theory of gender differences in cognitive ability rather than innate gender differences. The differences between males and females appear to be decreasing over time as the socialisation of children has shifted through the generations. The average IQ in economically developed countries has been increasing at a rate of about 2-3 IQ points every decade during the 20th century (Flynn, 1987; Lynn & Hampson, 1986). Raven et al. (1998) explained that the general improvement in IQ test scores has created a ceiling effect among more able adolescents and young adults. Raven et al. (1998) stated that to discriminate among more able individuals of these ages an extended version of the SPM has been developed. However, the skills and upbringing of boys and girls have become less differentiated (Feingold, 1988; Richardson, 1997), which could be due to improved environmental factors. There may be some evidence for this in that gender differences are the greatest between Black South Africans, and are less in Indian and Coloured populations, with almost no differences among Whites. Similar to that of cultural studies regarding the SPM, it appears that modern socialisation is creating less differentiation between males and females. However, the notion that gender differences are decreasing over time needs further investigation (Lynn, 2002). Another argument purposed by Kimura (1999) is that gender differences are influenced by early and current hormonal environments and thus intelligence has a genetic basis evolved by natural selection, but Abdel-Khalek & Lynn (2006) disputed this as consistent gender differences were not apparent in all cultures.

Ceiling and floor effects were noted from the analysis if the item difficulty. In both the Grade 8 and Grade 9 sample, the ceiling effects noted for the first few items on subtest A and the floor effects were noted for the last three to four items on subtest E. These findings were consistent with the claims that the items on the SPM increase in level of difficulty (Charmorro-Premuzic et al., 2005; Raven et al., 1998). It is not necessary to suggest that these items should be removed from the SPM as they do not provide information about individual differences. The purpose of the initial items is to orientate the test taker and the ending items show individual differences and exceptional ability (Raven et al., 1998).

The mean score for the Grade 8 sample was 42.3, while the mean score for the Grade 9 sample was 44.54, showing a higher score for the older group. Perhaps this improvement shows an improvement on stratum II with regard to Carroll's (1993) 3-stratum hierarchical theory of intelligence. This may show an improvement on broad intelligence such as fluid intelligence, broad visual perception or general psychomotor speed in the more mature group. It could also indicate an improvement in working memory (Carpenter et al., 1990). Perhaps with maturity, children improve certain skills incorporated in *g*, implying that intelligence is not a static phenomenon but rather something that can be improved through education and experience. The may also show an improvement in working memory as reported by Meo et al., (2006), Conway et al., (2002), Engel et al., (1990), Unsworth and Engel, (2005) etc.

In the past, studies have shown that certain ethnic groups have scored lower on the SPM than their western or white counterparts (Rushton, 2008; Owen, 1992; Israel, 2006) and general intelligence is the result of genetic makeup (Dickens et al., 2006; Nisbett, 2005; cited in Rushton, 2008; Pal et al., 1997). Other studies have disputed this (Lynn et al., 2005; Murray, 2003; Rushton, 2006; Carroll, 1997) by arguing that intelligence is the result of a process over a lifespan, thus arguing the environmental debate (Bock et al., 1986). The current study utilized a sample from diverse backgrounds. The language difference reported for the Grade 8 learners and not the Grade 9 learners, may show that intelligence is a developmental process. The current research may be evidence that one's environment may influence IQ scores as the Grade 9 learners who have had more exposure to equal schooling show no difference on the basis of language. Perhaps the EAL Grade 8 learners were unfamiliar with the constructs in the SPM as they may not have had equal opportunity to become familiar with the specific subject matter or processes required in order to answer a test item correctly (Eels et al., 1951; cited in Murphy et al., 1998). The Grade 9 EAL learners may also be more acculturated to their surroundings (Foxcroft et al., 2004) and urbanised (Freeman, 1984) having spent an extra year together in high school. Further research could investigate this phenomenon in order to note any difference on the SPM on the basis of language for younger and older samples. Perhaps the differences noted in the past are slowly disappearing as all children in South Africa are given equal opportunities to education (Ceci, 1991; Carroll, 1997).

Test bias in psychometric testing has long been a concern in the South African context. Many psychometric tests were developed for the western population and are associated with oppression and unfairness by different groups in South Africa (Foxcroft et al., 2004). Education in South Africa has historically been unequal between the racial groups creating privilege for white South Africans and disadvantage for the other racial groups (Mayekiso et al., 2007). Historically, the different groups in South Africa have performed unequally on the SPM favouring white South Africans (Rushton, 2008; Owen, 1992; Israel, 2006). Strategies, such as Feuerstein's Mediated Learning Experience (MLE) (Feuerstein, 1979; Skuy, Gewer, Osrin, Khunou, Fridjhon & Rushton, 2002; Rushton, Skuy & Fridjhon, 2002), need to be found to address the language issues in assessment in South Africa. Test should be available in all the official languages and regularly updated with appropriate and varied norms. There can be little doubt that the assessment of bilingual pupils, and those for who English is not their first language, creates a major problem for psychologists. The current research was concerned with whether the SPM is appropriate for the diverse South African population. This research provides evidence that the SPM should be administered with caution as a language difference was noted for the Grade 8 sample, though no difference was noted for the Grade 9 sample. The Australian and British norms also appeared to be inappropriate for both samples. This may imply that to completely disregard the SPM in South Africa may not be necessary, as long as the test measures the constructs in the same way for all the groups. Language and cultural differences on the SPM seem to be becoming increasingly less for the younger generation as they have had more equal opportunities than the older generations. The final percentiles should also not be used as fact as they may be lower than they should be. Perhaps the SPM could be used as a supplement to other test batteries. The individual's background, experience and access to education should also be considered before assessors make decisions (Carroll, 1997; Nettlebeck, 1998).

4.1. Limitations of the Research

It is important to note that there are several significant flaws in this research, possibly mitigating the findings outlined above.

Firstly, 161 of the learners were part of the Grade 8 sample as well as the Grade 9 sample. This repeated measure of this group in both samples may have affected the validity of the Grade 9 sample. The scores of these children may have been lower in Grade 9 had they not seen the test in both years, thus indicating a practice effect. Perhaps more significance between the language and gender groups would have been evident. However, these 161 learners only account for 39 % of the Grade 8 sample and 45% of the Grade 9 sample.

Secondly, the predictive validity of the data could have been calculated for both samples. If the SPM scores had correlated highly with the learners' school averages then the SPM could be seen as a good measure of general intelligence despite the language difference favouring the EFL learners in the Grade 8 data. If the SPM was deemed a good predictor of school performance with both samples then the language bias could be questioned. However, neither the SPM score nor the school averages are good indicators of potential but rather current functioning. If the EAL learners achieved lower averages and SPM scores that may also indicate a language difficulty rather than a lowered ability level.

There may be a threat to the external validity, the ability to generalise the research to other contexts and populations, as the sample was a convenience sample and was not randomly selected (Howell, 1999). However, because the school attended by the subjects was a government school that was formally a model C school, the children are varied across the race, culture and language groups. The school does not represent one particular population over another and therefore the sample was viewed as relatively representative of the wider population. However, it is unclear how well the sample, albeit a diverse sample, was representative of the population as the subjects were from one school.

In addition, although the subjects have all had at least 5 years of English as their medium of instruction, it is possible that some of the subjects may not have fully understood the requirements of the test during the instructions, or may have made the wrong interpretation about some aspects of the instructions. Furthermore, both the Grade 8 and Grade 9 sample represent two years of test administration implying that each group consisted of two administrations. Although every effort was made to standardise the administration procedure, the instructions given for each administration may not have been identical and therefore not equivalent across the groups. The dual administration and the one year interval between them may have caused a problem of diffusion. It is also possible that the subjects who were tested in both Grade 8 and 9 could have discussed the SPM outside of the testing setting, which may have affected their performance on the SPM.

4.2. Directions for Future Research

The findings of this research could be the base for further research regarding the SPM and other measures used by psychologist in South Africa. It may be useful to administer the SPM to representative groups of all school going ages, in order to investigate any language and gender differences across the age spectrum. In terms of language, today's children from English and non-English homes are exposed to English and the constructs of the western culture from a young age. More and more children are gaining a good education and therefore there is less disadvantage in psychometric testing for certain groups than in previous generations. It would be interesting to investigate if there is a language bias on the SPM for children with the same educational background. Perhaps at a certain age the language bias dissipates as children become increasingly familiar with the constructs in the test. Perhaps a developmental model will be noted in terms of language as it has shown with gender. A language difference may decrease with age rather than increase.

In addition, studies could be conducted at a number of different schools and with more learners from the South African population, to ascertain if findings are consistent across a broader population spectrum. South African norms on the SPM should be established as the British and Australian norms appeared to be inappropriate for the South African context. However, if norms were to be established, they should be updated regularly as the educational and social structures in South Africa are constantly shifting. It may also be useful to ascertain whether certain test items on the SPM are biased for the South African population and therefore develop a more appropriate version of the Raven's Progressive Matrices for the South African context. However, this too might fluctuate as the community problems such as poverty and level of education are gradually addressed with time in South Africa.

It may also be useful to measure the specific constructs and design features in each test item. Perhaps language differences should not be generalised over the whole SPM test, but rather the specific ability measured by each test item. Perhaps differences in proportions on each test item could indicate strengths and weaknesses for each group on specific abilities such as inductive ability (Rogers et al., 1994; cited in DeShon et al., 1995), fluid ability (Carrol, 1983; Cattell, 1973; cited in DeShon et al., 1995), working memory (Carpenter et al., 1990), deductive ability (Colberg, Nestor & Trattner, 1985; cited in DeShon et al., 1995), spatial ability (Ackerman & Kanfer, 1993; Hunt, 1972; cited in DeShon et al., 1995; Rohde et al., 2007), pattern perception (Dillon et al., 1981; Pande & Kothari, 1969; cited in DeShon et al., 1995) or non-verbal intelligence (Giles, 1964; Jensen, 1983; cited in DeShon et al., 1995).

In conclusion, this study adds to the body of research in South Africa regarding the usefulness of psychometric tests and in particular the Raven's Standard Progressive Matrices, a test that has not been standardised for the South African population. The performance on the SPM was measured for a group of Grade 8 and Grade 9 learners from a government school in Gauteng with regard to language and gender. A significant language difference was found in the Grade 8 sample only and no significant gender differences were noted for either grade. These findings raise concern regarding language bias in the SPM for high school children in South Africa, emphasizing that the test should be used with extreme caution in the South African context. Further studies should be conducted to confirm this finding to examine the particular ages and groups in which the language bias is common. The finding of no significant gender differences was consistent with research stating that males show an advantage on the SPM from age 15 to 16 onwards.

REFERENCE LIST

Abad, F. J., Colom, R., Rebollo, I. & Escorial, S. (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: evidence for bias. *Personality and Individual Differerences 36,* 1459-1470.

Abdel-Khalek, A. M. (1988). Egyptian results on the Standard Progressive matrices. *Personality and Individual Differences, 9(1),* 193-195.

Abdel-Khalek, A. M. & Lynn, R. (2006). Sex differences on the Standard Progressive Matrices and in educational attainment in Kuwait. *Personality and Individual Differences, 40,* 175-182.

Ackerman, P. L. (1988). Determinants of individual difference during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288-318.

Ackerman, P. L. (1990). A correlation analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, memory, and cognition, 16,* 883-901.

Ackerman, P. L. & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology, 78,* 413-432.

Amod, Z., Skuy, M., Lynn, S. & Fridjhon. (2000). Effectiveness of the Initial Assessment and Consultation Process (IAC) with Different Cultural Groups in South Africa. *International Journal of Special Education, 15(1),* 1-11.

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

Ankley, C. (1992). Sex differences in relative brain size: the mismeasure of women too? *Intelligence, 16,* 329-336.

Arendasy, M. & Sommer, M. (2003). The effect of different types of perceptual manipulation on the dimensionality of automatically generated figural matrices. *Intelligence, 33,* 307 – 324.

Arthur, W., & Woehr, D.J. (1993). A confirmatory factor analytic study examining the dimensionality of the Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement, 53,* 471-478.

Asmal, K. (2001). *Education White Paper 6: Special Needs Education Building an inclusive education and training system.* Pretoria: Elsen Directorate.

Ashman, A. F. & Conwah, N. F. (1991). An *introduction to cognitive education*. New York: Routledge.

Babcock, R. L. (1994). Analysis of adult age differences on the Raven's Advanced Progressive Matrices Test. *Psychology and Aging, 9,* 303 – 314.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (1990). *Human memory. Theory and practice.* Boston: Allyn and Bacon.

Beech, J. R. & Singleton, C. (1997). The psychological assessment or reading: theoretical issues and professional solutions. In J. R. Beech, S. and Singleton (Eds.). *The psychological assessment of reading.* (pp.1-26). London & New York: Routledge.

Bethell-Fox, C. E., Lohman, D. F. & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence, 8,* 205-238.

Bock, R.D., & Moore, E.G.J. (1986). *Advantage and disadvantage: A profile of American youth.* Hillsdale, NJ: Erlbaum.

Brody, N. (1992). *Intelligence*. San Diego, CA: Academic.

Caffarra, P., Vezzadini, G., Zonato, F. Copelli, S & Venneri, A. (2003). A normative study of a shorter version of Raven's progressive matrices 1938. *Neurological Sciences, 24,* 336-339.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press,

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence, 24(1),* 25-52.

Carpenter, P. A., Just, M. A. & Snell, P. (1990). What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven's Progressive Matrices Test. *Psychological Review, 97 (3),* 404 – 431.

Cattell, R. B. (1971). *Abilities: their structure, growth and action*. Boston: Houghton Mifflin.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27,* 703-722.

Chamorro-Presmuzic, T., Moutafi, J. & Furnham, A. (2005). The Relationship between Personality Traits, Subjectivity – assessed and fluid intelligence. *Personality and Individual Differences, 38,* 1517 – 1528.

Cole, M. & Means, B. (1981). *Comparative studies of how people think: An introduction.* Cambridge, MA: Harvard University Press.

Colom, R. & Abad, F. J. (2005). Advanced progressive matrices and sex differences: comment on Mackintosh & Bennet (2005). *Intelligence, 35,* 183-185.

Colom, R., Escorial, S. & Rebollo, I. (2004). Sex differences on the progressive matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences, 37,* 1289-1293.

Colom, R., & Garcia-Lopez, O. (2002). Sex differences in fluid intelligence among high-school gratuates. *Personality and Individual Differences, 32(3),* 445-451.

Colom, R., Garcia, L. F., Juan-Espinosa, M. & Abad, F. (2002). Null sex differences in general intelligence: Evidence from the WAIS-III. *Spanish Journal of Psychology, 5,* 29-35.

Colom, R. & Lynn, R. (2004). Testing the developmental theory of sex difference in intelligence on 12-18 year olds. *Personality and Individual Differences, 36,* 75-82.

Conway, A. R. A., Conwan, N., Bunting, M. F., Therriault, D. J. & Minkoff. S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed and general fluid intelligence. *Intelligence, 30,* 163 – 183.

Court, J. H. (1983). Sex differences in performance on Raven's Progressive Matrices: A review. *Alberta Journal of Educational Research, 29,* 54-74.

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

DeShon, R. P., Chan, D. & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for Multidimensional Performance. *Intelligence, 21,* 135 – 155.

Embretson, S. E. (2002). Generating abstract reasoning items with cognitive theory. In S. H. Irvine & P. C. Kyllonen (Eds.). *Item generation for test development*. New Jersey: Lawrence Erlbaum Associates.

Engel, R. W., Tuhdski, S. W., Laughlin, J. E. & Conway, A. R. A. (1990). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128,* 309 – 331.

Eysenck, H. J. (1994). A biological theory of intelligence. In D. K. Detterman (Ed.). *Current topics in human intelligence.* Norwood: NJ: Ablex.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43,* 95-103.

Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The Learning Potential Assessment Device, theory, instruments, and techniques*. Baltimore: University Park Press.

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

Flynn, J. R. (1998). Israeli military IQ tests: gender differences small; IQ gains large. *Journal of Biosocial Science, 30,* 541–553.

Freeman, M. C. (1984). *The effect of cultural variables on the Goodenough-Harris Drawing Test and the Standard Progressive Matrices*. Unpublished M.A. Dissertation, University of the Witwatersrand, Johannesburg.

Foxcroft, C., Paterson, H., le Roux, N & Herbst, D. (2004). *Psychological assessment in South Africa: A Needs Analysis. The test patterns and needs of psychological assessment practitioners.* Final Report.

Foxcroft, C. & Roodt, G. (2004). *An introduction to psychological assessment in the South African context.* Oxford: University Press.

Gregory, R. J. (2007). *Psychological Testing: History, Principles, and Application. (5th ed.).* Boston: Pearson.

Grieve, K. W. & Viljoen, S. (2000). An exploratory study of the use of the Austin maze in South Africa. *South African Journal of Psychology, 30,* 14-18.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. (3rd ed.). Mahwal, NJ: Erlbaum.

Howell, D. C. (1999). *Fundamental Statistics for the Behavioral Sciences (4th ed).* Pacific Grove: Brooks/Cole.

Irwing, P. & Lynn, R. (2005). Sex differences in means and variability on the Progressive Matrices in university students: A meta-analysis. *British Journal of Psychology, 96,* 505-524.

Israel, N. (2006). *Raven's Advanced Progressive Matrices within the South African context.* Unpublished masters thesis, University of the Witwatersrand, Johannesburg.

Jackson, D. N. & Rushton, J. P. (2006). Males have greater g: sex differences in general mental ability from 100,000 17- to 18- year-olds on the Scholastic Assessment Test. *Intelligence, 34,* 479-486.

Jensen, A, R. (1993). Spearman's Hypothesis Testes with Chronometric Information – Processing Tasks. *Intelligence, 17,* 47-77.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.

Jensen, M. R. & Feuerstein, R. (unknown date). Dynamic Assessment of Retarded Performers with the Learning Potential Assessment Device. To appear in *Dynamic Assessment: Foundations and Fundamentals.* New York: Guilford Press.

Kimura, D. (1999). *Sex and Cognition*. Cambridge, MA: MIT Press.

Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2004). Academic performance, career potential, creativity and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86,* 148-161.

Kyllonen, P. C. & Christall, R. E. (1990). Reasoning ability is (little more) than working memory capacity? *Intelligence, 14,* 389-433.

Logsdon, A. (2008). Nonverbal Intelligence - The Comprehensive Test of Nonverbal Intelligence. Retrieved from: http://learningdisabilities.about.com/od/intelligencetests/p/CTONI.htm on 16 October 2008.

Lynn, R. (1978). Ethnic and racial differences in intelligence: International comparisons. In R. T. Osborne, C. E. Noble, & N. Weyl (Eds.), *Human variation: The biopsychology of age, race, and sex* (pp. 261-286). New York: Academic Press.

Lynn, R. (1991). Race differences in intelligence: A global perspective. *Mankind Quarterly, 31,* 255-296.

Lynn. R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences, 17,* 257-271.

Lynn, R. (1997). Notes and shorter communications: Intelligence in Taiwan. *Personality and Individual Differences, 22(4),* 585-586.

Lynn, R. (1999). Sex differences in intelligence and brain size: A developmental theory. *Intelligence, 27,* 1-12.

Lynn, R. (2002). Sex differences on the progressive matrices among 15-16 year olds: some data from South Africa. *Personality and Individual Differences, 33,* 669-673.

Lynn, R., Backhoff, E. & Contreras-Nino, L. A. (2004). Sex differences on g, reasoning and visualisation tested by progressive matrices among 7-10 year olds: some normative data from Mexico. *Personality and Individual Differences, 36,* 779-787.

Lynn, R., Allik, J. & Irwing, P. (2004). Sex differences on the three factors identified in Raven's Standard Progressive Matrices. *Intelligence, 32,* 411-424.

Lynn, R & Hampson, S. L. (1986). The rise of national intelligence: evidence from Britain, Japan and the USA. *Personality and Individual Differences, 7,* 23-332.

Lynn, R & Irwing, P. (2004c). Sex differences on the progressive matrices: A meta-analysis. *Intelligenc,e 32,* 481-498.

Lynn, R., Raine, A. Venable, P. H. & Mednick, S. A. (2005). Sex differences in 3 year olds on the Boehm Test of Basic Concepts: Some data from Mauritius. *Personality and Individual Differences, 39,* 683-688.

MacArthur, R. S. & Elley, W. B. (1963). The reduction of socioeconomic bias in intelligence testing. *British Journal of Educational Psychology, 33,* 107-119.

Mackintosh. N. J. (1998). Sex differences and IQ. *Journal of Biosocial Science, 28,* 559-572.

Mackintosh, N. J. & Bennet, E. S. (2005). What do Raven's Matrices measure? An analysis of sex differences. *Intelligence, 33,* 663-674.

Matzen, L. B. L. V., van der Molen, M. & Dudink, A. C. M. (1994). Error Analysis of Raven's Test Performance. *Personality and Individual Differences, 16(3),* 433-445.

Mayekiso, T. & Tshemese, M. (2007). Contextual issues: Poverty. In N. Duncan, B. Bowman, A. Naidoo, J. Pillay & V. Roos. (Eds.), *Community Psychology: Analysis, context and action.* (pp. 150 - 165). Cape Town, S.A.: UCT Press.

MedTerms. (2008). Definition of Working Memory. Retrieved from: http://www.emedicinehealth.com/script/main/art.asp?articlekey=7143 on 16 October 2008.

Meo, M., Robert, M. J. & Marucci, F. S. (2006). Element salience as a predictor of item difficulty for Raven's Progressive Matrices. *Intelligence, 35,* 359-368.

Moran, A. P. (1986). The reliability and validity of Raven's Standard Progressive Matrices for Irish apprentices. *International Review of Applied Psychology, 35,* 533-538.

Murphy, K. R. & Davidshofer, C. O. (1998). *Psychological Testing: Principles and Applications (4th ed.).* New Jersey: Prentice-Hall.

Murray, C. (2003). *Human Accomplishment.* New York: Harper Collins.

Nettlebeck, T. (1998). Jensen's Chronometric Research: Neither Simple nor Sufficient But a Good Place to Start. *Intelligence, 26(3),* 233-241.

Nottcutt, B. (1949-50). The distribution of scores on Raven's Progressive Matrices Test. *British Journal of Psychology, 40,* 68-70.

Nyborg, H. (2005). Sex-related difference in general intelligence g, brain size, and social status. *Personality and Individual Differences, 39,* 497-509.

Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences, 13(2),* 149-159.

Owen, K & Chamberlain, J, C. (1996). Measurement and evaluation in psychology and education. In K. Owen & J. J. Taljaard (Eds.) *Handbook for the use of psychological and scholastic tests of the HSRC* (pp. 9-17). Pretoria: Human Sciences Research Counsel.

Pal, S., Shyman, R. & Singh, R. (1997). Genetic analysis of general intelligence 'g': A twin study. *Personality and Individual Differences, 22(5),* 779-780.

Primi, R. (2001). Complexity of geometric inductive reasoning tasks- contribution to the understanding of fluid intelligence. *Intelligence, 30,* 41 – 70.

Raven. J. C. (1965). *Advanced Progressive matrices, Set I and II.* London: H. K. Lewis.

Raven, J. (1989). The Raven Progressive Matrices: A Review of National Norming Studies and Ethnic and Socioeconomic Variation Within the United States. *Journal of Educational Measurement 26(1):* 1-16.

Raven, J. (1994). *Occupational User's Guide: Raven's Advanced Progressive Matrices & Mill Hill Vocabulary Scale.* Oxford: Oxford Psychologists Press.

Raven, J. (2003). *An overview or Raven's Progressive Matrices and Vocabulary Scales.* Oxford: OPP Limted.

Raven, J. C., Court, J. H. & Raven, J. (1977). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* London: H. K. Lewis.

Raven, J., Raven, J. C. & Court, J. H. (1998). *Section 3: Standard Progressive Matrices: Introducing the Parallel and Plus Versions of the Tests.* Oxford: Oxford Psychologists Press.

Raven, J., Raven., J, C. & Court, J, H. (2000) Standard Progressive Matrices: Oxford: Psychology Press.

Retief, A. (1988). *Method and Theory in Cross-Cultural Psychological Assessment*. Pretoria: Human Sciences Research Council.

Richardson, J. T. E. (1997). Conclusions. In P. J. Caplan, M. Crawford, J. S. Hyde & J. T. E. Richardson (Eds.). *Gender differences in cognition.* New York: Oxford University Press.

Rock, I. & Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American, 263,* 48-61.

Rohde, T. E. & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35,* 83-92.

Rosenthal, R. & Rosnow, R. (1991). *Essentials of Behavioural Research: Methods and Data Analysis (2nd ed.).* Boston: McGraw Hill.

Rushton, J. P. (1992). Cranial capacity related to sex, rank and race in a sample of 6,325 military personnel. *Intelligence, 16,* 401-413.

Rushton, J. P. (2006). In Defense of Disputed Study of Construct Validity from South Africa. *International Journal of Selection and Assessment. 14(4),* 381-384.

Rushton, J. P. (2008). Testing the genetic hypothesis of group mean IQ differences in South Africa: Racial admixture and cross-situational consistency. *Personality and Individual Differences, 44,* 786-776.

Rushton, J. P. & Skuy, M. (2000). Performance on Raven's Matrices by African and White University Students in South Africa. *Intelligence 28(4):* 251-265.

Rushton, J. P., Skuy, M. and Bons, T. A. (2004). Construct validity of Raven's advanced progressive matrices for African and Non-African engineering students in South Africa. *International Journal of Selection and Assessment, 12,* 220-229.

Rushton. J. P., Skuy, M. & Fridjhon, P. (2002). Jensen Effects among African, Indian, and White engineering students in South Africa on Raven's Standard Progressive Matrices. *Intelligence, 30,* 409-423.

Schweizer, K., Goldhammer, F., Rauch, W. & Moosbrugger, H. (2007). On the validity of Raven's matrices test: Does spatial ability contribute to performance? *Personality and Individual Differences, 43,* 1998-2010.

Skuy, M., Gewer, A., Osrin, Y., Khunou, D., Fridjhon, P. & Rushton, J. P. (2002). Effect of Mediated Learning Experience on Raven's matrices scores of African and non-African university students in South Africa. *Intelligence, 30,* 221-232.

Skuy, M., Hoffenburg, S., Visser, L. & Fridjhon, P. (1990). Temperament and cognitive modifiability of academically superior Black adolescents in South Africa. *International Journal of Disability, Development, and Education, 37,* 29-44.

Skuy, M., Schutte, E., Fridjhon, P. & O'Carroll, S. (2001). Suitability of the published neurological test norms for urban African secondary school students in South Africa. *Personality and Individual Differences, 30,* 1413-1425.

Skuy, M. & Shmukler, D. (1987). Effectiveness of the learning potential assessment device with Indian and Colored adolescents in South Africa. *International Journal of Special Education, 12,* 131-149.

Snyderman, S. & Rothman, S. (1988). *The IQ Controversy, the Media and Public Policy.* New Brumswick, Transection Books.

Spearman, C. (1923). *The nature of intelligence and the principles of cognition.* London: Macmillan.

Spearman, C. E. (1927). *The abilities of man, their nature and measurement*. New York: Macmillan.

Su¨ß, H. –M., Oberauer, K., Wittman, W. W., Wilhelm, O. & Schulze, R. (2002). Working-memory capacity explains reasoning ability – and a little bit more. *Intelligence, 30,* 261-288.

Sunberg, N. D. & Gonzales, L. R. (1981). Cross-cultural and cross-ethnic assessment: overview and issues. In P. McReynolds. (Ed.). *Advances in psychological assessment, Vol. 5*. San Francisco, CA: Jossey-Bass.

te Nijenhuis, J., van Vianen, A. E. M. & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence, 35,* 283 – 300.

Unsworth, N. & Engel, R. W. (2005). Working memory capacity and fluid abilities. Examining the correlation between operation span and Raven. *Intelligence, 33,* 67-81.

Valencia, R. R. (1984). Reliability of the Raven Coloured Progressive Matrices for Anglo and for Mexian-American Children. *Psychology in the schools, 21,* 49-52.

Van der Van, A. H. G. S. & Ellis, J. L. (2000). A Rasch Analysis of Raven's Standard Progressive Matrices. *Personality and Individual Differences, 29,* 45-64.

Vincent, K. R. (1991). Black/White IQ Differences: Does Age Make the Difference? *Journal of Clinical Psychology, 47(2),* 266-270.

Voyer, D., Voyer, S. & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration or critical variables. *Psychological Bulletin, 117,* 250-270.

APPENDIX 3a:
Biographical Questionnaire

BIOGRAPHICAL QUESTIONAIRE

NAME: ____________________

AGE: ____________________

GENDER: ________________

HOME LANGUAGE/S:

____

PLEASE FILL IN THE INFORMATION IN THE TABLE PROVIDED.

GRADE	SCHOOL	LANGUAGE OF INSTRUCTION
Grade One		
Grade Two		
Grade Three		
Grade Four		
Grade Five		
Grade Six		
Grade Seven		

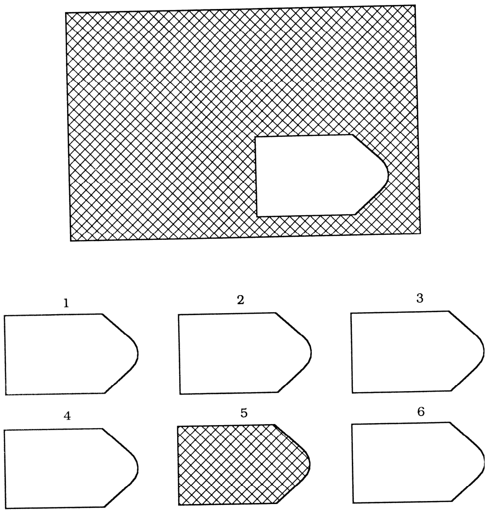
Do you speak in English to? (Please tick the appropriate box)

	Sometimes	Often	Never
Your Family			
Your Friends			

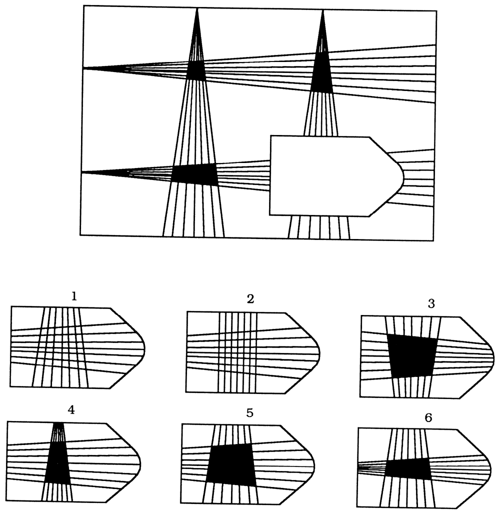
Thank you for taking the time to fill in this questionnaire. Your help is much appreciated.

APPENDIX 3b: Examples of items on the SPM

A2

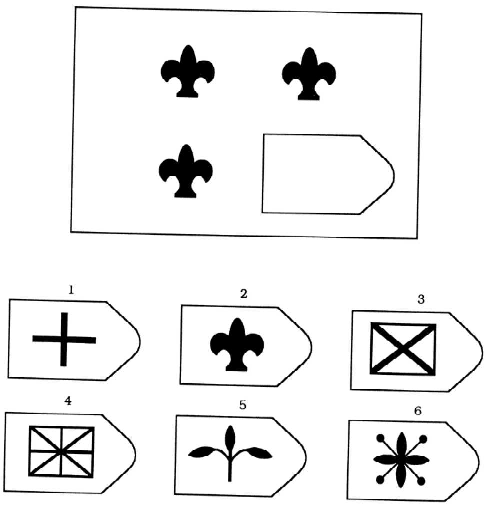


A12



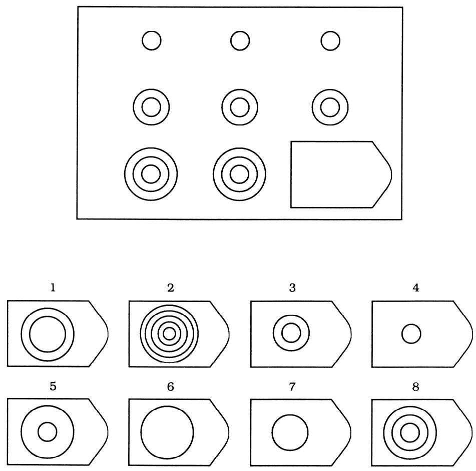
SET B

B1

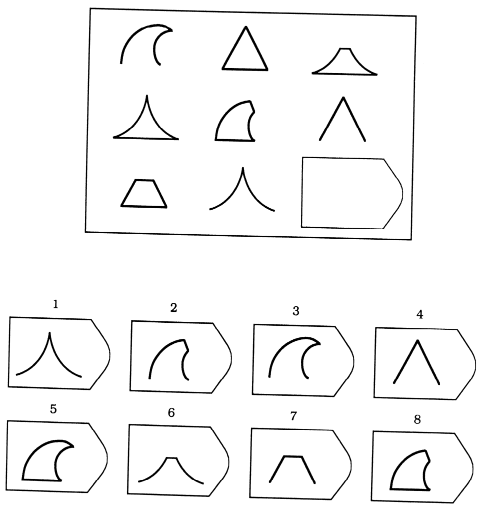


SET C

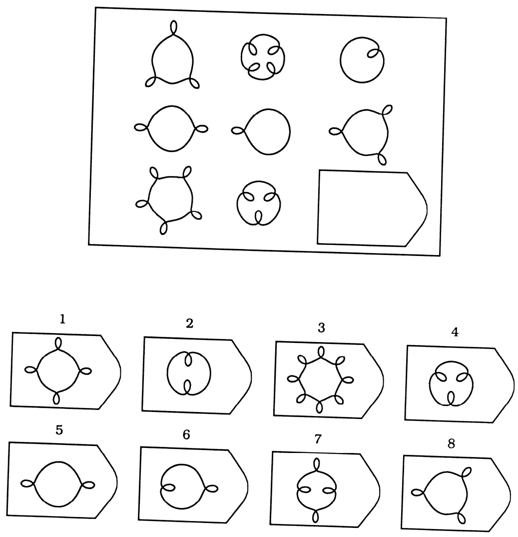
C1



D11



E12



APPENDIX 3c:
Consent Form

Consent Form

I, the parent/ guardian of ______________________________ consent to my son/daughter taking part in the reading research to be conducted at V [illegible] under the supervision of Dr Yvonne Broom from the University of the Witwatersrand.

I consent to the results of the Stanford Diagnostic Reading Test, a biographical questionnaire and academic marks being used for the purposes of this study.

I understand that:

- Participation in this research is voluntary.
- My son/daughter may withdraw from the study at any time.
- No information that may identify my son/daughter will be included in the research report, and all responses will remain confidential.

Name of Parent/ Guardian (Please print) ______________________________

Signed ______________________________ Date ______________________

APPENDIX 3d:
Assent Form

Assent Form

I ________________________________ agree to take part in the reading research to be conducted at \ )l under the supervision of Dr Yvonne Broom from the University of the Witwatersrand.

I agree that my results from the Stanford Diagnostic Reading Test, biographical questionnaire and academic marks may be used for the purposes of this study.

I understand that:

- Participation in this study is voluntary.
- I may withdraw from the study at any time.
- No information that may identify me will be included in the research report, and all responses will remain confidential.

Name of Learner (Please print) ________________________________

Signed ________________________________ Date _______________

APPENDIX 3e:
Letters to Parents

Information Sheet

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

School of Human and Community Development
Private Bag 3, Wits 2050, Johannesburg, South Africa
Tel: (011) 717-4500 Fax: (011) 717-4559

Dear Parent,

My name is ______________________, and I am conducting research for the purposes of obtaining a Masters Degree at the University of the Witwatersrand. My research focuses on the English reading abilities of first and second language learners who are being educated in English. South Africa has twelve official languages but the majority of learners choose to receive their Secondary School education in English, even though this is not their home language. The impact of learning in a language that is not your home language in not well understood in the South African context, and many learners may be at risk of underachievement due to poorly developed language ability. This research explores the level of reading abilities, particularly vocabulary and comprehension, of both first and second language learners, and its impact on their academic performance. The Principal, Governing Body and the Gauteng Department of Education have given permission for this research to be conducted at Willowmoore High School. We would like to invite your child to participate in this study.

Participation in this research will entail that your child completes a standard reading test, the Stanford Diagnostic Reading Test, which will be administered during school time by the teachers at . The test will be completed during the guidance lessons and the results will be collected from the guidance teacher by the researcher. In addition to the reading test your child will be asked to complete a short biographical questionnaire giving details of their English language experiences at home and during their Primary School education. Participation is voluntary, and no learner will be advantaged or disadvantaged in any way for choosing to participate or not participate in the study. Although your child will complete their name on the questionnaire to allow collation of this information with his/her academic marks, all responses will be kept confidential, and no information that could identify your child will be included in the research report. The results of individual learners' performance on the reading test will not be available to the learners or to the school.

If you choose to allow your child to participate in the study please complete the attached consent form and return it to the Class teacher at school. If you have any further queries about this research, please contact my supervisor, Dr Yvonne Broom ...ail at

Your participation in this study would be greatly appreciated. This research will contribute both to a larger body of knowledge on the level of English reading abilities of learners in South African schools. This will inform, and help us to develop, appropriate educational practices in our schools.

Kind Regards

APPENDIX 4a: Summary Statistics Grade 8

Table 13: Simple Descriptive Statistics Grade 8

Simple Descriptive Statistics									
Dep Var	**Gender**	**Home Lang**	**N**	**Mean**	**Std Dev**	**Median**	**Mode**	**Minimum**	**Maximum**
Tot A	Male	L1	12	11.31	0.76	12	12	9	12
		L2		11.12	1.19	11	12	5	12
	Female	L1		11.22	1.30	12	12	6	12
		L2		10.81	1.29	11	12	6	12
Tot B	Male	L1	12	10.39	1.47	10	11	6	12
		L2		10.12	1.82	11	11	3	12
	Female	L1		10.45	1.33	11	11	7	12
		L2		10.09	1.81	11	11	2	12
Tot C	Male	L1	12	8.55	1.88	8	9	3	11
		L2		8.54	1.82	9	9	0	12
	Female	L1		8.08	1.97	9	9	2	11
		L2		7.80	1.95	8	9	1	11
Tot D	Male	L1	12	8.64	1.95	9	9	1	12
		L2		8.54	1.84	9	10	1	11
	Female	L1		9.35	1.51	10	9	5	12
		L2		8.87	1.97	10	10	1	12
Tot E	Male	L1	12	4.73	2.44	4	3	0	12
		L2		3.79	2.22	4	2	0	10
	Female	L1		4.16	2.23	4	2	0	8
		L2		3.66	2.15	4	3	1	9

APPENDIX 4b:
Summary Statistics
Grade 9

Table 14: Simple Descriptive Statistics Grade 9

Simple Descriptive Statistics									
Dep Var	**Gender**	**Home Lang**	**N**	**Mean**	**Std Dev**	**Median**	**Mode**	**Minimum**	**Maximum**
Tot A	Male	L1	12	11.42	0.73	12	12	10	12
		L2		11.30	1.11	12	12	7	12
	Female	L1		10.92	1.05	11	11	8	12
		L2		10.95	1.41	11	12	5	12
Tot B	Male	L1	12	10.85	1.21	11	11	4	12
		L2		10.57	1.63	11	12	1	12
	Female	L1		10.68	1.47	11	12	4	12
		L2		10.73	1.23	11	12	5	12
Tot C	Male	L1	12	8.75	1.79	10	10	2	12
		L2		8.81	1.93	9	9	1	12
	Female	L1		8.72	1.93	9	9	1	12
		L2		8.52	1.60	9	9	4	12
Tot D	Male	L1	12	9.08	1.55	9	9	4	12
		L2		9.03	1.86	9	10	0	12
	Female	L1		9.32	1.87	10	10	1	12
		L2		9.36	1.68	10	10	4	12
Tot E	Male	L1	12	5.13	2.42	6	6	0	10
		L2		4.69	2.56	5	2	0	11
	Female	L1		5.02	2.60	5	6	0	10
		L2		4.50	2.22	4	6	0	9

APPENDIX 4c: Reliability Information

Correlation Analysis Grade 8

Table 15: Grade 8 Mean & Standard Deviation

Simple Statistics N = 413		
Variable	**Mean**	**Std Dev**
RAVENS TOTAL	42.31719	6.49037
TOTAL A	11.07022	1.16686
TOTAL B	10.21550	1.68072
TOTAL C	8.23002	1.92905
TOTAL D	8.77966	1.95517
TOTAL E	4.02179	2.26025

Table 16: Grade 8 Cronbach Coefficient Alpha

Cronbach Coefficient Alpha with Deleted Variable				
Deleted Variable	**Raw Variables**		**Standardized Variables**	
RAVENS TOTAL	1.000000	0.750324	0.993820	0.756335
TOTAL A	0.489105	0.779449	0.452080	0.861905
TOTAL B	0.609826	0.753395	0.568480	0.841164
TOTAL C	0.733607	0.727785	0.677770	0.820749
TOTAL D	0.681219	0.735075	0.618350	0.831964
TOTAL E	0.638608	0.733839	0.556734	0.843304

Correlation Analysis Grade 9

Table 17: Grade 9 Mean and Standard Deviation

Simple Statistics N = 375		
Variable	**Mean**	**Std Dev**
RAVENS TOTAL	44.53501	6.06954
TOTAL A	11.13445	1.12123
TOTAL B	10.70588	1.38827
TOTAL C	8.68908	1.80082
TOTAL D	9.20728	1.74229
TOTAL E	4.80392	2.44276

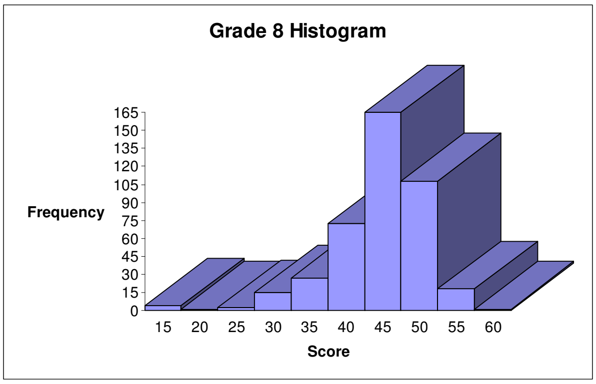
Table 18: Grade 9 Cronbach Coefficient Alpha

Cronbach Coefficient Alpha with Deleted Variable				
Deleted Variable	**Raw Variables**		**Standardized Variables**	
RAVENS TOTAL	0.999848	0.726037	0.990499	0.738871
TOTAL A	0.492340	0.771590	0.452472	0.848916
TOTAL B	0.553921	0.758240	0.512937	0.837748
TOTAL C	0.659158	0.731672	0.595396	0.822045
TOTAL D	0.656941	0.733944	0.597804	0.821578
TOTAL E	0.705064	0.703354	0.617930	0.817658

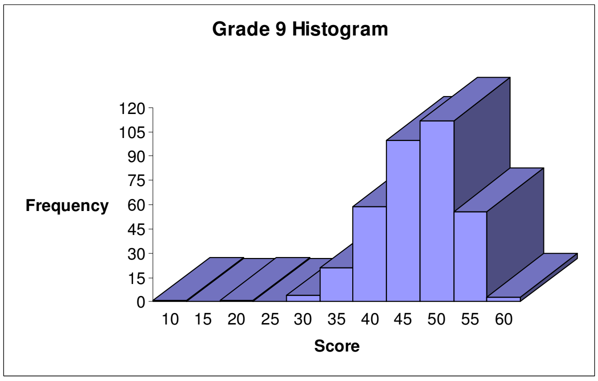
APPENDIX 4d:
Normality of the Data

Figure 1: Histograms: Grade 8 and Grade 9

Histogram Grade 8



Histogram Grade 9



The histograms showed the Grade 8 and Grade 9 data were normally distributed, although both data sets were slightly skewed to the left. In order to confirm the normal distribution of the data, a measure of central tendency was utilized reflecting where on the scale the distribution is centered, concentrating on the distance between the mean and the median (Howell, 1997).

Table 2a confirmed that the Grade 8 data is normally distributed as the mean (mean = 42.3) and median (median = 43) are sufficiently close in value (see Appendix 4a for further descriptive statistics). The means scores for each group were also sufficiently close to the median scores for the four groups. This allowed for parametric analysis of this Grade 8 data.

Table 2b confirmed that the Grade 9 data is normally distributed as the mean (mean = 44.5) and median (45) are sufficiently close in value. This allowed for parametric analysis of this Grade 9 data (see Appendix 4b for further descriptive statistics). The means scores for each group were also sufficiently close to the median scores for the four groups. This allowed for parametric analysis of this Grade 8 data.

Table 19: Kolmogorov-Smirnov Test of Normality for the SPM

Kolmogorov-Smirnov Goodness-of-Fit Test for Normal Distribution		
	Statistic (D)	**p – value**
SPM: Grade 8 Data	0.12661059	<0.001
SPM: Grade 9 Data	0.08084333	0.019

The results of the Kolmogorov-Smirnov Tests, as shown in Table 5, confirmed that the SPM data for both the Grade 8 ($p < 0.001$) and Grade 9 ($p = 0.019$) data sets were normally distributed.

APPENDIX 4e:
Australian and British Norms

Raven, Raven & Court (1998, p. 78)

Table SPM16

Standard Progressive Matrices

Smoothed 1986 Norms for Children and Young People in Australia in the Context of 1979 British Data

	Age in Years (Months)																	
	8½	9	9½	10	10½	11	11½	12	12½	13	13½	14	14½	15	15½	16	16½	17
	8(3) to 8(8)	8(9) to 9(2)	9(3) to 9(8)	9(9) to 10(2)	10(3) to 10(8)	10(9) to 11(2)	11(3) to 11(8)	11(9) to 12(2)	12(3) to 12(8)	12(9) to 13(2)	13(3) to 13(8)	13(9) to 14(2)	14(3) to 14(8)	14(9) to 15(2)	15(3) to 15(8)	15(9) to 16(2)	16(3) to 16(8)	16(9) to 17(2)
Percentile	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	UK AUS	AUS	AUS	AUS
95	42 44	44 46	46 48	48 49	49 50	50 51	51 51	52 52	53 53	54 53	54 54	55 54	56 55	57 55	57 56	56	57	58
90	40 42	42 44	44 46	46 47	47 48	48 49	49 49	50 50	51 51	52 51	53 52	54 52	54 53	55 54	55 54	55	56	57
75	36 39	38 39	41 42	42 43	43 44	44 45	45 46	46 47	47 48	49 49	49 50	50 50	50 51	51 51	51 52	52	53	54
50	31 32	33 34	36 36	38 38	39 39	40 41	41 42	41 43	42 44	43 45	44 45	45 46	46 47	47 47	47 48	48	49	50
25	22 22	25 25	28 28	32 31	33 33	34 35	36 36	37 38	38 39	39 40	41 41	42 41	42 42	42 42	42 43	44	45	45
10	17 13	17 15	19 20	23 24	27 27	29 29	31 31	31 33	32 34	33 35	35 36	36 36	36 37	36 37	36 38	39	40	40
5	13 11	14 12	14 14	17 16	22 19	24 22	25 25	26 27	27 29	28 30	29 31	30 32	33 33	33 33	33 34	35	36	37

APPENDIX 4f:
Post Hoc Analysis

Grade 8:

Table 20: Grade 8 Post Hoc Test

Dep Var	Significant Differences between Groups	P
Total Ravens	Male / EFL * Female / EAL	0.02
Total A	Male / EFL * Female / EAL	0.01
Total B	No significant differences	
Total C	Male / EFL * Female / EAL	0.01
	Male / EAL * Female / EAL	0.01
Total D	No significant differences	
Total E	Male / EFL * Male / EAL	0.01
	Male / EFL * Female / EAL	0.001

Grade 9:

Table 21: Grade 9 Post Hoc Test

Dep Var	Significant Differences between Groups	P
Total Ravens	No significant differences	
Total A	Male / EFL * Female / EFL	0.02
	Male / EFL * Female / EAL	0.02
Total B	No significant differences	
Total C	No significant differences	
Total D	No significant differences	
Total E	No significant differences	

APPENDIX 4g:
Grade 8 Item Difficulty

Table 22: Grade 8 Item Difficulty Index

Item Difficulty Index (*p*)	Male EFL	Male EAL	Female EFL	Female EAL
***p*A1**	1.0	1.0	1.0	1.0
***p*A2**	1.0	0.98	1.0	1.0
***p*A3**	1.0	0.98	1.0	0.99
***p*A4**	1.0	0.98	1.0	0.98
***p*A5**	0.99	0.99	0.98	1.0
***p*A6**	1.0	0.99	1.0	0.98
***p*A7**	0.98	0.93	0.90	0.92
***p*A8**	0.95	0.91	0.94	0.88
***p*A9**	1.0	0.95	0.96	0.97
***p*A10**	0.97	0.94	0.92	0.91
***p*A11**	0.85	0.83	0.82	0.71
***p*A12**	0.56	0.60	0.68	0.41
***p*B1**	1.0	0.98	0.98	0.99
***p*B2**	0.98	0.97	0.98	0.99
***p*B3**	1.0	0.95	1.0	0.97
***p*B4**	0.97	0.94	1.0	0.96
***p*B5**	0.94	0.94	0.96	0.86
***p*B6**	0.90	0.83	0.94	0.77
***p*B7**	0.79	0.71	0.74	0.65
***p*B8**	0.81	0.77	0.72	0.79
***p*B9**	0.81	0.81	0.96	0.83
***p*B10**	0.90	0.86	0.92	0.90
***p*B11**	0.77	0.83	0.76	0.84
***p*B12**	0.50	0.46	0.47	0.48
***p*C1**	0.97	0.98	0.98	0.95
***p*C2**	0.99	0.96	0.94	0.90
***p*C3**	0.91	0.92	0.88	0.90
***p*C4**	0.86	0.83	0.84	0.85
***p*C5**	0.92	0.88	0.90	0.81
***p*C6**	0.78	0.78	0.66	0.69
***p*C7**	0.79	0.89	0.84	0.80
***p*C8**	0.62	0.53	0.52	0.38
***p*C9**	0.80	0.81	0.90	0.75
***p*C10**	0.43	0.41	0.29	0.34
***p*C11**	0.39	0.45	0.27	0.33
***p*C12**	0.06	0.03	0.01	0.02
***p*D1**	0.97	0.99	1.0	0.99
***p*D2**	0.90	0.92	0.90	0.94
***p*D3**	0.90	0.84	0.98	0.88
***p*D4**	0.93	0.87	0.96	0.87
***p*D5**	0.93	0.94	0.92	0.95
***p*D6**	0.85	0.83	0.88	0.88
***p*D7**	0.74	0.73	0.84	0.72
***p*D8**	0.72	0.71	0.74	0.74
***p*D9**	0.67	0.70	0.80	0.70
***p*D10**	0.71	0.63	0.86	0.68
***p*D11**	0.18	0.18	0.29	0.32

***p*D12**	0.11	0.14	0.15	0.14
***p*E1**	0.75	0.82	0.78	0.68
***p*E2**	0.72	0.55	0.58	0.55
***p*E3**	0.69	0.60	0.58	0.58
***p*E4**	0.54	0.37	0.35	0.37
***p*E5**	0.51	0.36	0.49	0.39
***p*E6**	0.49	0.36	0.50	0.38
***p*E7**	0.32	0.26	0.35	0.27
***p*E8**	0.26	0.15	0.19	0.16
***p*E9**	0.11	0.16	0.17	0.10
***p*E10**	0.09	0.04	0.01	0.06
***p*E11**	0.08	0.03	0.03	0.02
***p*E12**	0.10	0.04	0.05	0.05

Figure 2: Grade 8: Item Difficulty GENDER

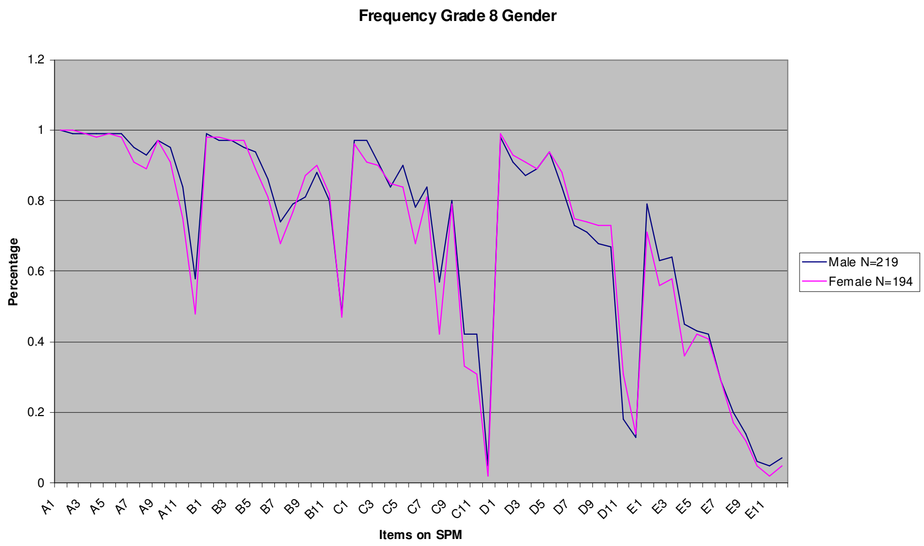
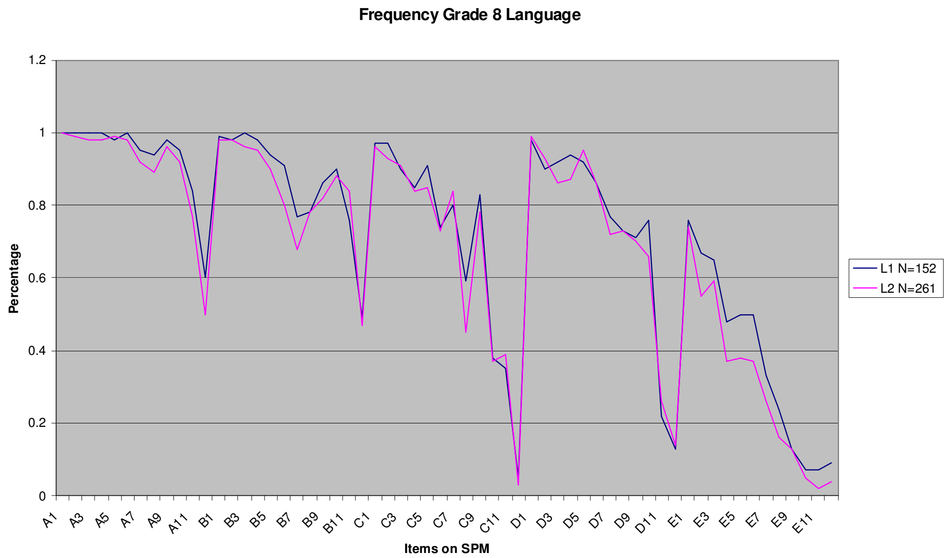


Figure 3: Grade 8: Item Difficulty LANGUAGE



APPENDIX 4h:
Grade 9 Item Difficulty

Table 23: Grade 9 Item Difficulty Index

Item Difficulty Index (*p*)	Male EFL	Male EAL	Female EFL	Female EAL
***p*A1**	1.0	1.0	1.0	0.99
***p*A2**	1.0	1.0	0.96	0.99
***p*A3**	1.0	0.98	0.98	0.98
***p*A4**	1.0	0.95	1.0	0.96
***p*A5**	1.0	1.0	0.98	0.97
***p*A6**	0.98	0.98	0.98	0.99
***p*A7**	0.96	0.95	0.96	0.88
***p*A8**	0.98	0.94	0.90	0.92
***p*A9**	1.0	0.96	0.96	0.92
***p*A10**	0.98	0.95	0.90	0.88
***p*A11**	0.84	0.86	0.82	0.84
***p*A12**	0.64	0.68	0.42	0.59
***p*B1**	1.0	0.98	1.0	0.99
***p*B2**	1.0	0.96	1.0	0.98
***p*B3**	1.0	0.98	1.0	0.97
***p*B4**	0.97	0.95	0.98	0.98
***p*B5**	0.98	0.96	0.92	0.95
***p*B6**	0.93	0.87	0.92	0.95
***p*B7**	0.84	0.84	0.81	0.71
***p*B8**	0.86	0.83	0.84	0.85
***p*B9**	0.89	0.86	0.87	0.91
***p*B10**	0.96	0.93	0.90	0.93
***p*B11**	0.82	0.80	0.85	0.88
***p*B12**	0.55	0.55	0.56	0.57
***p*C1**	0.97	0.95	0.97	0.99
***p*C2**	0.97	0.92	0.95	0.94
***p*C3**	0.93	0.94	0.93	0.96
***p*C4**	0.91	0.80	0.86	0.83
***p*C5**	0.89	0.92	0.96	0.89
***p*C6**	0.78	0.84	0.79	0.72
***p*C7**	0.92	0.92	0.92	0.88
***p*C8**	0.55	0.57	0.48	0.42
***p*C9**	0.82	0.88	0.90	0.82
***p*C10**	0.41	0.47	0.34	0.40
***p*C11**	0.40	0.48	0.42	0.52
***p*C12**	0.13	0.07	0.14	0.10
***p*D1**	0.97	0.96	0.97	0.99
***p*D2**	0.89	0.89	0.93	0.95
***p*D3**	0.88	0.85	0.96	0.94
***p*D4**	0.94	0.89	0.91	0.96
***p*D5**	0.94	0.96	0.91	0.97
***p*D6**	0.88	0.87	0.89	0.86
***p*D7**	0.84	0.78	0.80	0.85
***p*D8**	0.79	0.82	0.79	0.79
***p*D9**	0.72	0.84	0.79	0.81
***p*D10**	0.78	0.77	0.84	0.79
***p*D11**	0.17	0.24	0.32	0.23

***p*D12**	0.20	0.10	0.15	0.17
***p*E1**	0.78	0.74	0.78	0.80
***p*E2**	0.74	0.70	0.63	0.63
***p*E3**	0.77	0.68	0.67	0.71
***p*E4**	0.50	0.48	0.51	0.53
***p*E5**	0.65	0.47	0.70	0.46
***p*E6**	0.48	0.44	0.65	0.43
***p*E7**	0.43	0.35	0.41	0.32
***p*E8**	0.27	0.32	0.15	0.26
***p*E9**	0.16	0.21	0.18	0.15
***p*E10**	0.16	0.12	0.17	0.06
***p*E11**	0.06	0.06	0.08	0.01
***p*E12**	0.07	0.05	0.04	0.05

Figure 4: Grade 9: Item Difficulty GENDER

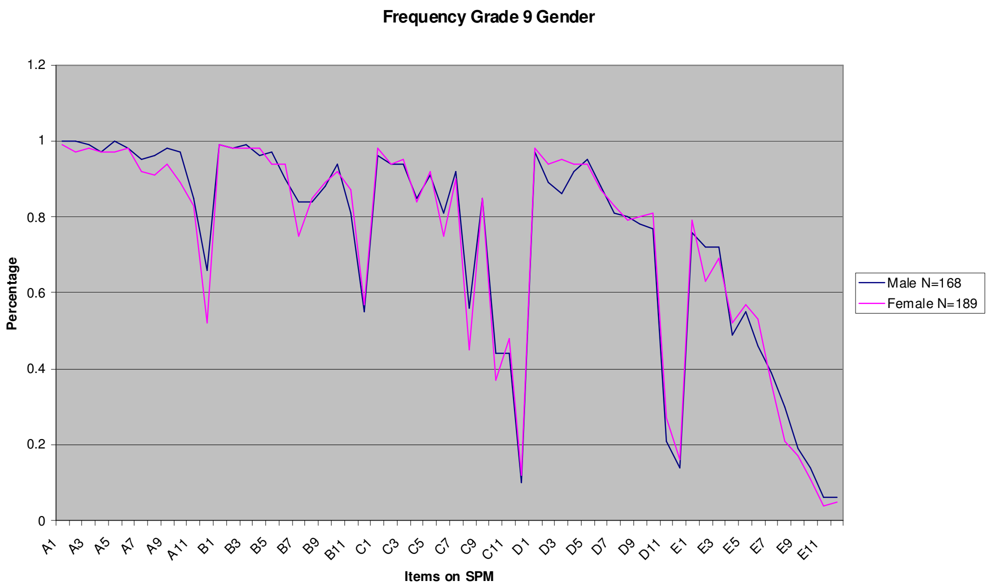
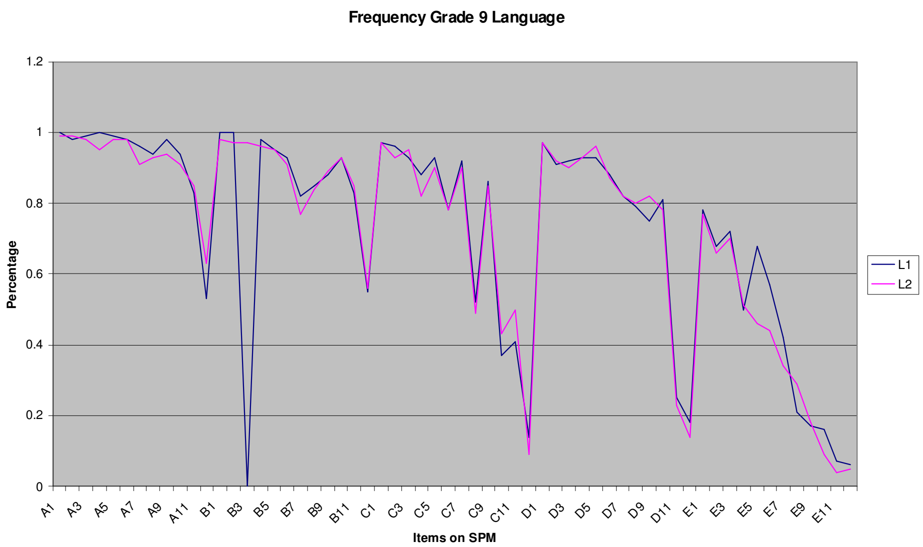


Figure 5: Grade 9: Item Difficulty LANGUAGE



APPENDIX 4i:
Matched Paired t-test

Table 24: Matched Paired t-test

	Diff	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Min	Max	DF	t Value	Pr > t
Gender 1 Home Lang 1	SPM TOT Gr8	35	-3.087	-1.486	0.1155	3.7704	4.6613	6.1072	0.7879	-16	5	34	-1.89	0.0679
Gender 1 Home Lang 2	–	43	-3.86	-2.186	-0.512	4.4847	5.4391	6.9131	0.8295	-11	11	42	-2.64	0.0117
Gender 2 Home Lang 1		22	-2.862	-0.545	1.7715	4.0204	5.2256	7.4678	1.1141	-12	8	21	-0.49	0.6295
Gender 2 Home Lang 2	SPM TOT Gr9	61	-4.728	-3.197	-1.666	5.0729	5.9772	7.2769	0.7653	-27	8	60	-4.18	<.0001
All vars		161	-3.048	-2.193	-1.337	4.9562	5.4983	6.1746	0.4333	-27	11	160	-5.06	<.0001
Gender 1 Home Lang 1	TOT A Gr 8	35	-0.367	-0.086	0.1952	0.6616	0.8179	1.0716	0.1382	-1	2	34	-0.62	0.5394
Gender 1 Home Lang 2	–	43	-0.277	0.1163	0.5091	1.0524	1.2763	1.6222	0.1946	-2	3	42	0.60	0.5534
Gender 2 Home Lang 1		22	-0.265	0.2273	0.7193	0.8538	1.1098	1.5859	0.2366	-2	2	21	0.96	0.3477
Gender 2 Home Lang 2	TOT A Gr9	61	-0.686-	-0.393	-0.1	0.97	1.1442	1.393	0.1465	-4	2	60	-2.69	0.0094

	Diff	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Min	Max	DF	t Value	Pr > t
Gender 1 Home Lang 1	TOT D Gr 8	35	-0.443	0.1143	0.6718	1.3127	1.6228	2.2162	0.2743	-4	4	34	0.42	**0.6796**
Gender 1 Home Lang 2	–	43	-0.866	-0.395	0.0754	1.2613	1.5297	1.9443	0.2333	-4	3	42	-1.69	**0.0975**
Gender 2 Home Lang 1		22	-0.595	0.2727	1.1401	1.505	1.9562	2.7956	0.4171	-5	4	21	0.65	**0.5203**
Gender 2 Home Lang 2	TOT D Gr 9	61	-1.209	-0.656	-0.103	1.8327	2.1594	2.6289	0.2765	-8	4	60	-2.37	**0.0209**
Gender 1 Home Lang 1	TOT E Gr 8	35	-1.397	-0.571	0.2545	1.9449	2.4045	3.1504	0.4064	-8	5	34	-1.41	**0.1688**
Gender 1 Home Lang 2	–	43	-1.937	-1.116	-0.296	2.1979	2.6656	3.388	0.4065	-7	5	42	-2.75	**0.0088**
Gender 2 Home Lang 1		22	-1.772	-0.818	0.1361	1.6558	2.1522	3.0757	0.4589	-6	4	21	-1.78	**0.0890**
Gender 2 Home Lang 2	TOT E Gr 9	61	-1.369	-0.787	-0.205	1.9299	2.2739	2.7683	0.2911	-6	5	60	-2.70	**0.0089**