Evaluating the Stanford Diagnostic Reading Test as a measure of reading comprehension and vocabulary for South African Grade Eight learners.

Deborah Diane Catto (329154)

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 is my own independent work, and had not been presented for any other degree at any academic institution, or published in any form.

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

_____

Deborah Diane Catto
(329154)

Acknowledgements

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

My supervisor, Dr Yvonne Broom, for all her expertise, guidance, encouragement and support. Your warmth and sense of calm were greatly appreciated

Mr Mike Greyling for his statistical assistance and expertise

Mr Peter Fridjhon for his statistical direction and expertise

Ms Nicky Israel for her assistance in interpreting the data

My fellow M. Eds, Jolene, Bianca, Shalya, Lara, Gail, Anwynne, Viviana, and Aileen, for all your encouragement and support through this process, which through sharing together, was made easier

My parents and my brother, for your unfailing support, encouragement and love.

Abstract

One of the needs of assessment practitioners in South Africa is for the adaptation or revision of current psychological and educational tests to justify their use in the South African context. The South African Department of Education called for the re-evaluation of all standardised tests used in the South African context (Department of Education, 1997). The Stanford Diagnostic Reading Test (SDRT) is an English language proficiency test that is useful in measuring reading comprehension. The fact that English is the language of instruction in most South African schools makes the SDRT a valuable measure for South African assessment practitioners. This study evaluated the SDRT as a measure of vocabulary and reading comprehension for South African Grade Eight learners on the basis of home language and gender. Two subtests (Auditory Vocabulary and Reading Comprehension) of the Brown level of the SDRT were administered to three consecutive years of Grade Eight learners. Of 631 subjects, all of who had received at least five years of English medium education, 279 were English First Language (EFL) learners and 352 were English Additional Language (EAL) learners. There were 316 male subjects and 315 female subjects. The results showed that the EAL learners performed significantly below the EFL learners on both the Vocabulary and Reading Comprehension subtests of the SDRT. There was also a significant difference in performance between male and female learners on the Comprehension subtest but not on the Vocabulary subtest. Both the Vocabulary and the Reading Comprehension subtests demonstrated adequate internal consistency. However, the validity of these subtests was questioned because a number of items appeared to demonstrate bias against EAL learners.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Literature Review

The current use of psychological and educational tests in the South African school context is a contentious issue. The justification for using such tests on South African learners is being questioned, both by government policies and researchers in the area. Inclusive education policies in South African schools stress the recognition of individual learning needs and differences as well as practice of non-discrimination in all contexts (Department of Education, 2001). Assessment practices in South African schools are required to embrace such principles to ensure that assessment is ethical and in the best interests of the child. This presents a challenge to both the South African educator and the assessment practitioner.

In 1997 the National Commission on Special Needs in Education and Training (NCSNET) and the National Committee on Education Support Services (NCESS) submitted a report entitled "Quality Education for All: Overcoming barriers to learning and development". In this report they specified that the application of standardised tests to learners should only be justified if "there is reasonable guarantee that no learners subjected to the assessment will be disadvantaged in any way" (Department of Education, 1997, p. 85). Furthermore, the NCSNET and NCESS (Department of Education, 1997, p. 85) recommended that, "urgent attention should be given to the re-evaluation of all standardised tests". Therefore, the South African Department of Education demanded research in the area of psychological and educational assessment including the analysis of standardised tests that are used in South Africa for educational assessment of learners.

Foxcroft, Paterson, Roux & Herbst (2004) conducted a national study in South Africa that identified the needs of assessment practitioners and which tests they used. Major concerns of assessment practitioners included the lack of psychological tests adapted for South Africa's diverse context; inappropriate norms; and a space for the development of uniquely South African tests. Furthermore, this study highlighted that both national and international tests, currently used by South African assessment practitioners, need to be "urgently adapted, revised or updated" (Foxcroft et al. 2004, p.7). This significant and extensive study provides the foundation for research into the use and adaptation of international tests used in South Africa.

The Stanford Diagnostic Reading Test (SDRT) is listed as a commonly used formal reading test (Lerner, 1993). It is widely used because it is one of the few formal reading tests that measure different levels of comprehension (Karlsen & Gardner, 1986). In the South African school context, the assessment of different levels of reading comprehension is crucial for supporting all learners. By Grade Eight, most learners have mastered decoding skills but many struggle with reading comprehension, particularly English Additional Language (EAL) learners (Pretorius, 2000). Difficulty in reading comprehension is a barrier to learning in terms of hindering understanding of English textbooks and other texts (Pretorius, 2000). Therefore, a test that identifies a learner's strengths and areas for growth in terms of different levels of reading comprehension such as the SDRT is a valuable asset to the assessment practitioner. The two subtests of the SDRT that assist in understanding a learner's ability to make meaning of an English text such as a school textbook are the Auditory Vocabulary and Reading Comprehension subtests. So these subtests were explored in the study.

An important part of interpreting scores is their comparison to a normative sample so that the assessor can understand the learner's ability relative to their age-related peers (Wolfaardt, 2001). The SDRT is an American developed test that has been standardised for the U.S. population. This means that the normal distribution of scores is for that specific population. Practitioners using the SDRT to assess South African learners need to be cautious when interpreting scores relative to American norms. Consequently, research that determines the South African reliability and validity of the SDRT would be valuable in assisting practitioners to use the SDRT responsibly.

This study examines the difference in performance between English First Language (EFL) learners and English Additional Language (EAL) learners and male and female learners on two subtests of the SDRT. This report will review the relevant literature that provides the theoretical and practical context of this study. An overview of psychological and educational assessment will be presented followed by a discussion of assessment in South Africa. Test translation and test adaptation will be examined within the South African context. Lastly, theory and research related to the SDRT, vocabulary development and reading comprehension, and cultural and gender bias will be discussed.

The professional practice of educational psychologists includes the use of psychological and educational assessment for the purpose of identifying a child's strengths and educational needs, tracking development or growth, and providing information that is useful for the

purpose of making decisions about interventions (Foxcroft & Roodt, 2001). Early notions of assessment date back to Biblical times and 2200 B.C. China (Foxcroft & Roodt, 2001; Gregory, 2007). However, it is only over the last 100 years with the development of intelligence tests and the scientific method that testing has taken its place firmly in the realm of psychology and education (Foxcroft, Roodt & Abrahams, 2001; Kriegler & Skuy, 1996). The advent of standardisation and the development of norms that allowed comparison between individuals or groups first occurred in 1905 (Foxcroft et al., 2001). The First World War initiated the process of group testing and the use of multiple-choice as a test format. Significantly, there was criticism at this time that questioned the appropriateness of intelligence tests for illiterate people (Foxcroft et al., 2001). Therefore, critical evaluation of the applicability of standardised tests for all individuals has been a concern since the origins of the tests themselves.

The root word of assessment is the Latin word “'assidere', which means to sit beside” (Kriegler & Skuy, 1996, p. 113). This definition of assessment suggests a closeness to the person being assessed, a sharing of experience, and a conscious effort to understand that person's perspective. It opposes the idea that assessment is only reliable and valid if it is entirely objective and relative to a normed population (Kriegler & Skuy, 1996). Furthermore, this original perspective highlights the importance of understanding the person's unique social and cultural context in ensuring that the picture created through assessment is as accurate and meaningful as possible. This definition implies that the purpose of assessment is to understand the individual being observed. Responsible assessment practitioners strive to gain the best possible understanding of the individual. Such an undertaking requires assessment of different aspects of functioning (attention, motor, cognitive, verbal, non-verbal, scholastic achievement, personality-related) over a period of time so that patterns of functioning may be identified and is achieved by gathering information from a number of sources such as teachers, parents, and other professionals as well as through the use of assessment measures or tests (Foxcroft & Roodt, 2001).

Tests are generally conceived as standardised measures used for a specific objective such as measuring an individual’s performance on various tasks or as a measure of a sample of behaviour (Anastasi, 1982; Foxcroft & Roodt, 2001). The term 'standardised' or ‘standardisation’ refers to uniformity across settings in reference to both administration and scoring of the test (Anastasi, 1982; Gregory, 2007). Therefore, a test is also defined as “a

standardized procedure for sampling behaviour and describing it with categories or scores" (Gregory, 2007. p. 2).

Owen (1998, p. 13) argues that the most important function of a psychological test is "to measure inter- and intra-individual differences, in other words differences between individuals as well as differences in the various individuals themselves". Tests are referred to as being either norm-referenced or criterion-referenced tests. Norm-referenced tests are "tests that use a well-defined population of persons for their interpretive framework" (Gregory, 2007, p. 4) while criterion-referenced tests are "tests that measure what a person can do rather than comparing the results to the performance levels of others" (Gregory, 2007. p. 4).

Foxcroft and Roodt (2001) prefer to use the term "assessment measures" instead of "tests" to incorporate a broader definition of assessment. However, most literature still uses the term 'test' and, therefore, these terms will be used interchangeably. Importantly, they highlight the relationship between assessment measures and "the context in which they are designed" (Foxcroft & Roodt, 2001. p. 6), specifically that:

> A measure is usually designed in a certain context (society, culture) for a specific purpose, and the normative information used to interpret test performance is limited to the characteristics of the normative sample. Consequently, the appropriateness of an assessment measure for an individual, group, or organisation from another context, culture, or society cannot be assumed without an investigation into possible test bias and without strong consideration being given to adapting and renorming the measure (Foxcroft & Roodt, 2001, p. 6).

Essentially, this statement highlights the fact that every test, whether norm-referenced or criterion-referenced, is context-bound and, therefore, is only appropriate for use with the population from whence it came. Nevertheless, test designers generally strive to create tests or assessment measures that can be applicable across cultures.

Culture is commonly defined as the "shared patterns of beliefs and knowledge by which people order their perceptions and experiences and in terms of which they act" (Weber & Tardif, 1991). Therefore, cross-cultural testing involves extricating from a test those features that may disadvantage people whose beliefs, knowledge, perceptions and experiences are different to those who designed the test (Anastasi, 1982). Intelligence tests have been

criticised for their reliance on language and verbal ability, for example, and therefore, non-verbal measures such as the Raven Progressive Matrices (RPM) were developed (Foxcroft et al., 2001). The value attached to speed is another factor that varies across culture and, therefore, cross-cultural tests attempt to limit the influence of time on performance (Anastasi, 1982). Lastly, test-content is frequently very culture-specific and thus, the content of cross-cultural tests needs to be considered carefully to make certain that people from different cultures have the experiential background required to understand the content (Anastasi, 1982).

Furthermore, psychometrists require that assessment measures be reliable and valid. An assessment measure is described as reliable when it is able to measure an aspect of behaviour consistently and as valid when it accurately measures what it claims to measure (Murphy & Davidshofer, 2001). Both reliability and validity can be compromised when a test is used in a context other than the one in which it was developed.

1.1. Assessment in South Africa

In the South African context it is essential that assessment practitioners carefully examine the instruments with which they evaluate a person's abilities, particularly when the measures that they use have been constructed and standardised in another country with another population. Is it really plausible to use a test that was conceptualised and validated in an entirely different context and apply it to South Africa where factors such as culture, language and socio-economic status play such a crucial role in a person's perspective and understanding?

The South African context demands a unique approach to educational and psychological assessment. This is due to the fact that the environment in which assessment in South Africa developed was "characterised by the unequal distribution of resources based on racial categories" (Foxcroft et al., 2001, p.22). The early development of psycho-educational testing in South Africa was, therefore, marked by the fact that standardisation was only for the White population and, therefore, tests were highly biased in favour of one group, White South Africans, as well as the fact that results from these tests were used to perpetuate Apartheid philosophy and policy (Foxcroft et al., 2001).

Significantly, S. Biesheuvel challenged the research, upon which White intellectual superiority was supposedly indicated, in 1943 in a book entitled 'African Intelligence'.

According to Foxcroft et al. (2001, p.23), Biesheuvel questioned the "cultural appropriateness of Western-type intelligence tests for blacks [*sic*]". Thus, the question of whether a certain test is meaningful for all population groups in South Africa was raised even at a time when the use of such standardised measures on such a disadvantaged population was clearly arbitrary.

Today, many South African assessment practitioners are still working from a deficit-model, assuming that if a child performs poorly in a test that it is some problem within the child (Kriegler & Skuy, 1996). However, factors such as lack of formal education; lack of resources; contextual factors; and language barriers have a significant impact on how an individual child performs in a test. South Africa does not have valid psychological and educational tests that fully accommodate all these factors. Indeed, according to Kriegler and Skuy (1996), the usefulness of Western standardised tests is highly questionable.

In Foxcroft et al.'s (2004) study, assessment practitioners including clinical, counselling, educational, industrial and research psychologists as well as psychometrists, psychotechnicians, registered counsellors, psychology students and interns were questioned about their needs related to psychological and educational tests. Amongst the issues raised was a need for "more cross-culturally appropriate tests" (Foxcroft et al., 2004, p. 36) and "a need for more training with regards to the adaptation, application, and interpretation of tests for cross-cultural purposes" (Foxcroft et al., 2004, p. 36). A limitation of this study was that a relatively small sample of the population responded to the survey (N = 881, response rate = 13%). However, the authors argue that the sample reflected the characteristics of South African psychologists and that the findings were corroborated by the findings of two other studies (one focus group study and one individual interview study) conducted by the same authors.

Importantly, this study also showed that "South African psychologists perceive testing to be central to their professional functioning" (Foxcroft et al., 2004, p. 131). Some of the reasons included the argument that psychological tests are objective and well researched; that they provide rich information in a short period of time; and that they provide standardised information that allows for comparison between individuals (Foxcroft et al., 2004). However, the participants also highlighted the fact that psychological tests add value only if they are valid, reliable and cross- culturally applicable (Foxcroft et al., 2004).

Language is just one dimension to South African diversity but it is an important one and one that is recognised internationally. Robert Gregory (2007, p. 38) cites Valdés and Figueroá (1994, p. 172):

> When a bilingual individual confronts a monolingual test, developed by a monolingual individual, and standardized and normed on a monolingual population, both the test-taker and the test are asked to do something that they cannot. The bilingual test taker cannot perform like a monolingual. The monolingual test can't "measure" in the other language.

This implies an inherent disadvantage for the EAL learner being measured on a test developed for an EFL learner. In this way language barriers play a critical role in determining whether a test is appropriate for English Additional Language learners (Foxcroft et al., 2004).

South Africa has eleven official languages. Seventy four percent of the population speak an African language as their first language, 15% of South Africa's population speak Afrikaans, and 9% of the population speak English as their first language (Broom, 2001). Significantly, despite the statistics and due to South Africa's political past, two languages, Afrikaans and English, have been given precedence in South Africa (Martin, 1997). During the apartheid regime under the guise of promoting multilingualism and in the form of the Bantu Education Act of 1953, the apartheid government effectively isolated Black learners from Afrikaans and English education systems (Martin, 1997). Following the Bantu Education Act, the Department for Native Affairs introduced the '50:50 policy', which insisted on teaching being divided between English and Afrikaans and non-academic subjects being taught in the learners' first language (Martin, 1997). This policy together with a more stringent language policy during the 1970s, led to the 1976 Soweto Uprising, in which Black learners rioted in protest of educational oppression and the language policies that had been enforced (Broom, 2001).

Past educational policies in South Africa created a scenario in which the majority of South African learners 'choose' to learn in their second, third or fourth language, hence the use of the term English as an Additional Language (EAL) instead of English as a Second Language (ESL). A number of studies suggest that "bilingualism can positively affect both intellectual and linguistic progress" (Cummins, 1991, p. 164). The linguistic advantage of bilingual speakers is attributed to the fact that they have mastered two languages allowing for more

experience in analysing meanings (Cummins, 1991). The relationship between bilingualism and cognitive development is based upon the person achieving a higher threshold of bilingual competence (Ben-Zeev, 1984). However, it also appears to be dependent upon the person experiencing an additive form of bilingualism. Additive bilingualism refers to a situation where learning a second language is at no cost to a person's first language and the first language is either dominant in the environment or perceived as prestigious (Ben-Zeev, 1984; Cummins, 1991). In contrast, subtractive bilingualism occurs when there is no educational support of a person's first language and learning a second language is to the detriment of the development of both languages or the loss of the first language (Ben-Zeev, 1984; Cummins, 1991). For the majority of South African learners their educational experience has been characterised by a form of subtractive bilingualism as a result of the country's political past. Therefore, it follows that studies that have demonstrated the positive effects of bilingualism may possibly only be generalised to White bilingual South African learners because their education reinforced the privilege of their languages in the community and thus provided them with the cognitive skills to achieve in education (Heugh, 1995b as cited in Broom, 2001).

Current educational policies in South Africa strive to promote an additive form of bilingualism by teaching in the majority first language up unto Grade Three and then in English from Grade Four onwards (Martin, 1997). However, this policy appears to be difficult to implement because many parents desire their children to be taught in English from the outset due to the high status of English both locally and internationally (Martin, 1997). This is in spite of the fact that numerous research studies have shown that submerging learners in a second language at the expense of the first language does not necessarily promote second language proficiency (Cummins & Mulcahy, 1978; Legaretta, 1979; Leyba, 1978; Skutnabb-Kangas & Toukomaa, 1976, Troike, 1978, as cited in Cummins, 1981; Rosier & Farella, 1976). Furthermore, studies have shown that an additive form of bilingualism may promote cognitive and academic advantage over monolinguals (Cummins, 1981).

Cummins (1981) distinguishes between basic interpersonal communicative skills (BICS) and cognitive/ academic language proficiency (CALP). He argues that BICS "involves processing language within a meaningful interpersonal context in which word meaning is supported by many situational and paralinguistic cues" (Cummins, 1981, p. 23) whereas

CALP involves higher order literacy skills which is context-reduced (Cummins, 1981; Kessler, 1984). This distinction is important, especially within the South African context and with reference to assessment, because Cummins (1984, as cited in Cummins 1991) reported that psychologists often fail to take into account the difference between these two types of proficiency when testing EAL learners. Although, these learners may appear to be fluent in English, that is, they have mastered BICS, they may not have adequate CALP. Cummins (1980, as cited in Miller, 1984) suggests that an adequate level of CALP is only reached after five to seven years of residence in a country or experience with the language, after the age of six years old.

In this study, the EAL learners have at least five years of English medium education, and therefore, should have developed adequate CALP. Furthermore, these learners are generally South African citizens and thus it is likely that they have had experience of English since birth. However, the subtractive nature of many of these learners' bilingualism may have prevented some of these learners from becoming proficient in English cognitive or academic language. The developmental interdependence hypothesis accounts for why EAL learners who have been exposed to English for a number of years may not be proficient in either their first or additional language. This theory postulates that second or additional language acquisition is influenced by the level of first language proficiency (Miller, 1984). The hypothesis extrapolates that intensive exposure to a second or additional language whilst the first language is underdeveloped may hinder the further development of the first language and subsequently, the acquisition of the second language (Miller, 1984). It can be inferred based upon the above discussion that this might be the case for a number of South African learners.

Therefore, South African assessment practitioners need to be keenly aware of the issues of language such as bilingualism, the relationship between bilingualism and academic achievement and the difference between basic interpersonal communicative skills and cognitive or academic language proficiency. Such issues are crucial to understanding assessment in the South African context, and particularly the use of English language tests. However, discussion of such issues also leads to recognition of the need for English language proficiency measures in South Africa to assist in gauging the level of competence of South African learners in the English language.

Culture is another important element that comes into play when a test is used on a different population to one on which the test was developed (Beech & Singleton, 1997; Foxcroft et al., 2004). Most tests are intrinsically culture or context-bound and, therefore, beg the question of whether standardised tests are ever truly bias-free. Indeed, Anastasi (1982, p. 342) refers to a "continuum of cultural differentials" in reference to the influence of culture for the individual, implying that it is impossible to be completely isolated from the influence of culture. South Africa is a multicultural society, which implies that it is a society that consists of a number of different perspectives, beliefs, practices and experiences. There appears to be a common perception that due to South Africa's diversity and its socio-political past, it presents a unique context. Furthermore, it seems that there is an urgency to make assessment measures in South Africa relevant for South Africans (Foxcroft et al., 2004).

Culture and context are two terms that often appear to be used interchangeably. However, Duquette (1991) argues that there are subtle differences between the terms, specifically in reference to the relationship between culture or context and language acquisition. He refers to context as the surrounding environment or circumstances that influence behaviour and culture as "group-determined behaviour" (Duquette, 1991, p. 122). He also cites a definition of context in relation to language acquisition that incorporates culture:

> Contextual refers in this case to the inner as well as the outer context, not only to situational features, but also to linguistic and emotional factors that are often strongly influenced by the attitudes and values of the speaker's cultural environment. (Rivers, 1983, as cited in Duquette, 1991, p. 122).

Based upon the above definition, culture may be described as being part of context. Therefore, in discussing cultural factors that may influence test performance, the context of the test-takers is explored. However, there is more to context than culture. The context of a country, region or community also includes geographical and climatic features, language, socio-economic factors, and common commercial and recreational activities. The South African context is characterised by geographical and climatic features such as mountainous regions and flatlands, wet summers and dry winters or dry summers and wet winters, in different regions. Many South African children have never seen snow. The South African context is also marked by the great divide between the haves and the have-nots. Some South African children may have the privilege of travelling overseas, whilst many do not travel beyond their own town. In terms of recreational activities, socio-economic factors, again,

play a significant role. The majority of South African learners may not have the means to go the cinema or watch television, for example. However, there are some who are accustomed to enjoying skiing or snowboarding during the holidays.

Thus, the diversity of experiences in South Africa makes it difficult to define a singular context of South African learners. Similarly, the multicultural nature of South African society begs the question of whether different assessment measures or norms should be developed for different groups, both cultural and socio-economic, for example. And yet, this is neither a practical nor an economic alternative (Owen, 1998). Furthermore, such a move could also be described as a type of discrimination. There are some psychologists who have advocated that assessment measures either be modified or newly created for specific cultural groups (Helms, 1992 & Davidson, 1995, as cited in Owen, 1998). In answer to such proposals, Dyck (1996, as cited in Owen, 1998, p. 81) states that "such a racially specific approach to assessment is based on inappropriate racial stereotyping, [and] a confounding of cultural (categorical) variables with individual (continuous) variables". Thus, there exists the potential for discrimination even under the pretext of making a measure more culturally appropriate.

Shuttleworth-Jordan (1996, p. 96) goes even further and appeals against a form of nihilism in South Africa, that assumes that "all tests in common usage on westernised populations should be abandoned and new culturally relevant and appropriately standardised tests should be designed". She cautions that there is often "erroneous exaggeration of cultural effects, which fails to take into account the acculturation process" (Shuttleworth-Jordan, 1996, p. 96). She argues that in the South African context, people are at different stages of urbanisation, westernisation, and literacy. She further highlights that differences in test performance are frequently attributed to cultural differences when it is in fact differences in level of education that have led to the discrepancy (Ardila, 1995, as cited in Shuttleworth-Jordan, 1996). However, she does note that culture "dictates what is and what is not relevant, and provides models for ways of thinking, acting and feeling" (Ardila, 1995, as cited in Shuttleworth-Jordan, 1996, p. 97-98) and that this does influence test performance.

Significantly, Shuttleworth-Jordan (1995a, as cited in Shuttleworth-Jordan, 1996) conducted a comparative research study that compared the performance of African first language and English first language students of equivalent university education in a range of non-verbal and verbal tests (SA WAIS: Digit Span and Digit Symbol, Wechsler Memory Scale: Paired Associates, Logical Memory and Visual Reproduction, Verbal Fluency - Words in one

minute, The Trail Making Test, and the Finger Tapping Test). The results indicated that although the African first language students scored consistently below the English first language students, the difference between the mean scores of the two groups was within one standard deviation and, therefore, had little clinical significance, according to Shuttleworth-Jordan (1996). In explaining the poorer performance of the African first language students, she argues that the long-term educational history of the African first language students may have provided fewer opportunities for quality education in comparison to the educational history of the English first language students. This, together with earlier socialisation experiences, is the most probable cause of the difference in performance between the two groups. Shuttleworth-Jordan (1996) lastly, warns that misattribution of lower test scores to socio-cultural factors risks the failure to identify real impairments in cognitive or educational functioning for this population.

Therefore, it is the responsibility of educational psychologists to seek ways in which assessment can be meaningful and true for every individual. Shuttleworth-Jordan (1996) recommends that one way of doing this is by using standardised tests, even with urban African language speakers studying in English, as part of a holistic assessment approach. In addition, she advises that South African research should focus on "the modification and standardization of existing test sources for use with urbanized black populations" (Shuttleworth-Jordan, 1996, p. 107). Other suggestions include the use of criterion-referenced tests as these are "specific to the child's own progress and do not involve comparison with other pupils" (Beech & Singleton, 1997, p. 17). However, criticism can also be made of criterion-referenced tests in that the criteria itself is often norm-related (Beech & Singleton, 1997).

Finding ways to assess every individual fairly is not a simple or easy task. Standardised tests have significant value in terms of being well-researched, valid and reliable measures of very defined constructs. Therefore, it is essential that researchers find ways to adapt tests so that they become more appropriate for the context but retain their reliability and validity. Many of the psychological and educational tests that are used in South Africa are internationally developed tests, mostly from North America or the United Kingdom (Foxcroft et al., 2004). Research into these tests requires evaluation of their South African validity, reliability and cross-cultural bias (Foxcroft et al., 2004).

The preceding discussion highlights both the need for an awareness of the effects of socio-cultural and contextual factors on test development and performance as well as recognition of the fact that cultural effects are frequently given too much credit for the differences in performance between different groups of people. In the South African context, factors such as quality of education and socialisation processes undoubtedly have an effect on test performance amongst the different population groups. However, this is not sufficient justification to abandon the use of standardised tests in South Africa. The alternative is the translation or adaptation of these tests for use in South Africa.

1.2. Test Translation and Test Adaptation

The history of test adaptation in South Africa spans a period of time from colonial inquisition to apartheid educational policies to the 'new' South Africa where government policies attempt to right the wrongs of the past. In Retief's (1988, p.14) 'Method and theory in cross-cultural psychological assessment', he provides an overview of "adaptability testing, in the broadest sense of the word" as he calls it, which has involved both an attempt at adapting the Black population to the demands of westernised tests and adapting westernised tests to the needs of the Black population. This debate, about the role of western psychology in the African context, is beyond the scope of this study. However, test adaptation for the purpose of using standardised tests in the South African context is key in the current practice of psychological and educational assessment.

It is useful to differentiate between test translation and test adaptation. Test translation means translating a test from one language into another without changing the meaning of the test (Kanjee, 2001). In contrast, test adaptation aims to retain the meaning of the original test but also strives to make the test "more applicable to a certain context while using the same language" (Kanjee, 2001, p. 87). In other words, adapting a test involves changing words, examples, and contexts in order to make the test appropriate for a different context (Kanjee, 2001). Test adaptation is preferable to test translation because test translation can possibly over-simplify the process and in so doing compromise the validity of the test (Hambleton, 1996). For example, Shuttleworth-Jordan (1996) notes that in the Xhosa translation of Hemp's (1989) memory test for children there are many words that were unfamiliar to Eastern Cape Xhosa-speaking children, living in an urban area. Although this was based upon informal clinical observations, she argues that it indicates that the English version of

this test is more relevant to that population. Therefore, test translation can be complicated because it may depend upon assumptions about a particular population or language group.

However, test adaptation is also a complex process, with a poor test adaptation possibly leading to a significant change in scores (Hambleton, 1996). In Hambleton (1996), the importance of maintaining the meaning of the constructs being measured is highlighted. However, he also emphasises taking consideration of whether "the construct itself is meaningful in the target group" (Hambleton, 1996, p. 13). Therefore, test adaptation does not merely involve changing words or contexts but demands an evaluation of whether a specific construct has value to a certain group of people. For example, in some cultures reflection and thoughtfulness is valued in comparison to the Western value placed on the ability to respond quickly (Hambleton, 1996; Kanjee, 2001). Such differences in value would impact on the way in which a test would be approached.

Thus the challenge in South Africa, in terms of test adaptation, is defining both the context and the values of the general South African population. This is an almost impossible task due to the multiplicity of contexts, cultures and the varying degrees of westernisation in South Africa. Therefore, research efforts should endeavour to identify traits and values common to all South Africans so that tests may be adapted to meet the needs of the general South African population.

Test adaptation is also viewed in conjunction with norming (Foxcroft et al., 2004). Norms are misinterpreted when they are based on a population that is not representative of the population for which the norms are being used (Murphy & Davidshofer, 1998). South Africa does not have norms that have been developed for the multicultural population, and yet, assessment practitioners believe that the development of South African norms is necessary in ensuring that tests produce valid and reliable results (Foxcroft et al., 2004). Skuy, Schutte, Fridjhon, & O' Carroll (2001) confirmed the need for appropriate norms for the South African population in their study that compared the performance of urban African students on neuropsychological tests with the corresponding American norms and provided the beginnings of an appropriate set of norms for the urban African population.

In summary, tests need to accurately measure what they set out to measure and demonstrate consistency in measurement, specifically for the South African population. However, once again, practicality and expenditure concerns prevent the development of South African

norms. Consequently, it is essential in the South African context to explore the possibilities of adapting international assessment measures to reduce bias and ensure that assessment is meaningful, reliable and valid for the population that is being measured (Kanjee, 2001).

1.3. The Stanford Diagnostic Reading Test

The Psychological Corporation in the USA first published the SDRT in 1966. In 1974 the 1st edition was supplemented with a higher level for use in high schools and colleges (Karlsen & Gardner, 1986). The latest edition, the Fourth Edition was published in 1996. The SDRT is an English Language Proficiency test. Its target population is English First Language speakers and its Grade levels range from Grade One to Grade Twelve. It can also be used to assess low-achieving college or university students.

The fact that the test is an English Language Proficiency test immediately begs the question of why it would be used with EAL learners. In South Africa, there is a national move towards English as the primary medium of education despite the fact that the majority of the population are not English First Language speakers (Pluddemann, 1999). Therefore, a large proportion of EAL learners attend English-medium schools and are required to learn through the medium of English using English textbooks. In fact, most textbooks are either available in English or Afrikaans (Pluddeman, 1999). As a result, it is imperative that educators are able to reliably determine how well or how poorly learners understand what it is that they are reading in English textbooks. The SDRT is a useful instrument through which to do this.

The SDRT is a favoured test for two primary reasons. Firstly, it makes use of diagnostic criteria, which provides detailed information about different reading skills (Karlsen & Gardner, 1986). Secondly, the SDRT "places more emphasis on the low achiever" (Karlsen & Gardner, 1986, p. 3). This allows for a more accurate assessment of these learners' abilities. The authors of the SDRT believe that a diagnostic reading test should be responsive to all learners (Karlsen & Gardner, 1986). Such a perspective is accommodated well within the South African inclusive education framework and provides good reason for the use of the SDRT in South Africa and an investigation into its South African reliability and validity.

Furthermore, in terms of supporting EAL learners in having to learn through the use of English textbooks, knowledge about their reading comprehension skills is indispensable. Reading comprehension involves understanding meaning on three levels: word, sentence, and

text (Bowyer-Crane & Snowling, 2005). The two subtests in the SDRT that provide information about these different levels of comprehension are the Auditory Vocabulary and Reading Comprehension subtests.

For the purposes of this research study, only these subtests were administered. The Auditory Vocabulary subtest provides information about language proficiency, in other words, understanding the meaning of words, and uses words that are representative of the different parts of speech that "were sampled from three general content areas: reading and literature, mathematics and science, and social sciences and the arts" (Karlsen & Gardner, 1986, p. 6). The Reading Comprehension subtest assesses literal and inferential comprehension using three types of comprehension passages: textual, functional, and recreational. The texts represent a variety of content that is appropriate for the age range.

1.3.1. Vocabulary and reading comprehension

Word recognition, understanding what the word means and then understanding the meaning of the entire sentence is a simple way of describing some of the processes involved in reading. Therefore, assessment of a learner's vocabulary and reading comprehension is crucial in understanding their level of reading development. Difficulties in both the development of vocabulary and in reading comprehension are attributed to multiple factors. The issues highlighted here are those, which may be relevant to understand the reading difficulties of South African learners.

A number of studies have shown that print exposure may account for differences in vocabulary development (Cunningham & Stanovich, 1991, Stanovich & Cunningham, 1992, 1993, as cited in Stanovich et al., 1996). However, the authors note that their findings may be viewed as examples of the "Matthew effects" where the rich get richer and the poor get poorer, meaning that children who are already fluent readers will read more and therefore, expand their vocabulary and in turn improve their comprehension and their academic skills (Stanovich et al., 1996). This coincides with Reitsma and Verhoeven's (1998, p. 4) statement that "literacy problems can also be related to children's social and economic background", in terms of exposure to reading material.

Similarly, underdeveloped vocabulary has been found to correlate with biligualism (Ben-Zeev, 1972, 1975; Rosenblum and Pinker, 1983, as cited in Ben-Zeev, 1984). Ben-Zeev

(1984, p. 56) explains that bilinguals may have fewer experiences in hearing any particular word because "they must share their conversations between two language systems". In addition, studies have shown that a person's knowledge of vocabulary is strongly correlated with his or her ability to comprehend what he or she is reading (Anderson & Freebody, 1983, as cited in Stanovich et al., 1996). A specific difficulty that may become a barrier to bilingual readers' comprehension was identified in a study that indicated that second language readers may make predictions about language structure based on cognitive strategies used in their first language, thus leading to confusion (Cowan & Sarmed, 1976, as cited in Ben-Zeev, 1984).

Vocabulary scores have also been found to be closely related to Spearman's general intelligence factor and the Peabody Vocabulary Test is used as a measure of general intelligence (Ben-Zeev, 1984). Therefore, reduced vocabulary may indicate lower cognitive functioning. However, vocabulary deficits are also suggestive of reading difficulties or language barriers (Ben-Zeev, 1984).

According to Oakhill and Cain (1998), research has shown that difficulties in reading comprehension may be the result of a number of factors. These include word recognition that is not sufficiently smooth and automatic, underdeveloped syntactic knowledge that hinders a person at the level of the sentence, memory difficulties, and difficulties with inference and integration of information (Oakhill & Cain, 1998). Furthermore, to understand the meaning of a sentence a person draws upon their knowledge of the world (Gough, Hoover, & Peterson, 1996). Therefore, comprehending a text involves a number of complex processes that lead to a mental representation of what the text describes. Oakhill and Cain (1998) highlight the fact that integration of the meaning of sentences and the inferences made from the meaning of these sentences requires some degree of identification of the main ideas of the text. The authors offer the example, adapted from Charniak (1972, as cited in Oakhill & Cain, 1998), of the following text:

> Jane was invited to Jack's birthday party. She wondered if he would like a kite. She went to her room and shook her piggy bank. It made no sound.

They argue that to understand this text requires knowledge about social conventions associated with birthday parties and with saving money. The concept of a piggy bank, for instance, may be completely foreign to some people. However, without this knowledge the

text would not make sense. The text also requires the reader to make inferences. For example, the reader must infer that Jane does not have the money to buy Jack a kite because there was no money in her piggy bank (It made no sound). This example provides a basis for the analysis of the texts used in the SDRT in reference to the context of South African learners.

1.3.2. Possibilities for Cultural Bias in the SDRT

For the purpose of this research, a definition of cultural bias that is found in Murphy & Davidshofer (1998, p.313) will be used:

> Cultural bias refers to differences in the extent to which the child being tested has had the opportunity to know and become familiar with the specific subject matter or specific process required by the test item.

Accordingly, cultural bias in tests used in South Africa could exist in both the content of items as well as their format. Careful consideration of both these elements is required in determining how culturally biased a certain test could be.

The item format, in other words, the way in which questions are framed, of the SDRT is multiple choice. However, Kanjee (2001, p. 89) states that “South African students are more familiar with essay type questions while those in the USA are more familiar with multiple-choice questions”. Thus, this suggests that the format of the SDRT itself could be a disadvantage to some South African learners. However, this problem can be accommodated quite simply through the use of practice items that orientate the learner with the format. The impact of cultural background on reading comprehension is not as easily remedied.

A person's cultural background has a significant influence on their reading comprehension. Incongruency between the learner's context upon which interpretation is based and the text presented leads to difficulty in comprehending the meaning of the text (Machet, 1991). Furthermore, the emotions that the content of the text elicits in a person will also colour the way in which the text is understood (Machet, 1991). Therefore, the selection of material for the reading comprehension subtest is crucial to how well a learner will understand the meaning of the test. In the South African context, material should be relevant to the general context of the South African population.

There is a growing interest in cross-cultural research (Hambleton, 1996). However, there is a significant gap in literature, generally, around test adaptation for different cultural groups and, specifically, on the cultural bias of the SDRT. In addition, based upon the preceding literature and particularly the findings of Foxcroft et al's (2004) national study on the needs of South African assessment practitioners, it is clear that the usefulness of standardised tests is questionable because many of these tests are not appropriate for the diverse South African context. There is a decisive call for research into the adaptation of current tests as well as the development of new measures tailored for the unique South African context. The SDRT is a test that is used in South Africa but "can only be justified if it can be demonstrated that the information gathered will be useful to the child from whom it is obtained" (Kriegler & Skuy, 1996, p. 116).

1.3.3. Possibilities for Gender Bias

Murphy and Davidshofer (1998, p. 301) state that "bias is said to exist when a test makes systematic errors in measurement or prediction". Gender bias thus refers to a test or item that fails to accurately measure a construct or predict achievement for either males or females. In a reading comprehension test, gender bias may exist in the content of the passages that have been chosen as well as the types of texts that are used. However, the difference in performance between males and females on reading tests cannot simply be accounted for by potential bias in test content or structure. Issues such as interest in reading, enjoyment of reading and an environment that promotes reading are all factors that contribute to reading achievement. Nevertheless, the research studies, highlighted below, have found that females generally perform better on reading tests, particularly reading comprehension tests.

A research study conducted by Rutter, Caspi, Fergusson, Horwood, Goodman, Maughan, Moffitt, & Carroll (2004) has shown that reading difficulties are far more common in boys than in girls. This UK and New Zealand-based study summarised the findings of previous research into gender differences in reading difficulties and aimed to generate new empirical evidence of the findings through four epidemiological studies that targeted the general population (Rutter et al., 2004). The strengths of these studies included the large size of the samples, the high participation rates and the representative nature of the samples for the general population, and lastly, the fact that the results were based on standardised testing. The report by the authors provides significant evidence that boys tend to display higher rates of reading difficulties (Rutter et al., 2004).

In another study involving 199, 097 15 year olds from 43 countries, girls outperformed boys on the reading comprehension test in every country. Girls were also more likely to be adequate readers. One of the variables that was found to partly account for this gender difference was interest in reading, both individual and school-based (Chiu & McBride-Chang, 2006). This study also highlighted the contextual factors that influence reading development and gender differences in reading achievement including socio-economic status of the school a child attends, number of books at home, reading enjoyment, and the influence of a peer culture of enjoyment of reading, which the authors argue may be facilitated by a higher concentration of girls in a school (Chiu & McBride-Chang, 2006).

In addition, a different study involving 66 students aged eight and nine found that girls had a more positive attitude towards reading than boys (Lynch, 2002). This study examined the relationship between parents' reading beliefs, children's reader self-perceptions, and their reading achievement in terms of gender (Lynch, 2002). Significant relationships were found between parents' self-efficacy beliefs and children's self-perception of their reading, specifically that mothers had stronger beliefs "in their ability to help improve boys' reading achievement" (Lynch, 2002, p. 54) and between children's own perceptions about their reading and reading achievement.

According to Pomplun & Omar (2001), item bias research (Rosser, 1989, Scheuneman, 1983, Wendler & Carlton, 1987, Zwick & Ercikan, 1988, as cited in Pomplun & Omar, 2001) has shown that items with content concerning war or the military may disadvantage women. However, their study, which involved a sample of 2000 tenth graders and confirmatory factor analysis, found invariance across both genders for a reading comprehension test using an expository passage pertaining to the American Civil War (Pomplun & Omar, 2001). Thus, the authors suggested the continued use of passages about war in comprehension assessments.

Interest in the content of texts used in reading comprehension assessments is another factor that has been considered in view of identifying gender bias. In Bray and Barron's (2003) study, involving 19, 735 students from Grade Four to Grade Eight and 98 different reading passages, they found that girls demonstrated greater interest in texts with predominantly female characters while boys' interest appeared to decrease with poetry. However, they caution against removing such content and emphasise rather awareness about the potential

effect of interest on test scores. In addition, the authors recommend that test developers do not use a single reading passage or passages relating to a single theme as they did find a small but significant relationship between interest and comprehension, which could influence test outcomes (Bray & Barron, 2003).

Therefore, gender differences in overall performance on the Vocabulary and Comprehension subtests of the SDRT and in performance on individual items would be interesting to note. Furthermore, the validity of the measure includes inspection of whether certain items of the SDRT are biased in favour of either sex. Bias could be incorporated through test content in terms of its greater interest and relevance to one sex (Neale, 1989).

1.3.4. The SDRT: Previous research

Although a number of studies, both internationally and locally, have used the SDRT as an instrument (Israel, 2006; Lathy, 2006; Ping-Ha & Chi-Ting, 2000; Pretorius, 2000), there seem to be few studies evaluating the SDRT itself. There is one study that has examined the difference in performance of university students on the Blue level of the SDRT (refer to second para. in Chapter 2.3. for description of the Blue level) and also evaluated the SDRT as a predictor of academic achievement (Andor, 2006). Her findings indicate that second language university students performed significantly below first language university students on all the subtests of the Blue level of the SDRT, including vocabulary and reading comprehension (Andor, 2006). She also found that the SDRT was not a good predictor of academic achievement for either first language or second language university students. However, the author does note that there were a number of limitations to her study such as the effects of confounding variables, a small number of participants and greater English First Language than English Additional Language representation (Andor, 2006). Although the current study acknowledges a number of confounding variables, the sample size is large and adequately representative of the population.

The rationale for the current study emerged out of a need recognised at a number of levels. Firstly, policies presented by government have directed such a course of action. Secondly, an extensive national research study demonstrated that assessment practitioners in South Africa identify the need to re-evaluate the standardised tests that are currently used in South Africa and develop new tests that are fully applicable to South Africa. There is a gap in the literature

in terms of research into the appropriateness of the use of the SDRT in the South African context. And lastly, the SDRT is a useful test in assessing reading comprehension, however, its suitability for the South African context needs to be scrutinised.

This study aimed to investigate using the SDRT for the assessment of English auditory vocabulary and reading comprehension in the South African context by determining the reliability and validity of the SDRT for South African Grade Eight learners. The difference in the performance of English First Language (EFL) learners and English as Additional Language (EAL) learners and Male and Female learners on the SDRT was also explored. Finally, a comparison was made between English First Language (EFL) learners and English Additional Language (EAL) learners' performance and Male and Female performance on individual items in the Vocabulary and Comprehension sub-tests of the SDRT. The underlying reasons for any differences or similarities between these groups were considered.

The following three research questions were investigated:

- How reliable and valid are the Vocabulary and Comprehension subtests of the SDRT for South African Grade Eight learners, specifically for EFL learners, EAL learners, Male learners, and Female learners?
- Is there a difference in performance on the Vocabulary and Comprehension subtests of the SDRT between English First Language (EFL) learners and English Additional Language (EAL) learners and Male and Female learners?
- Is there a difference in performance on individual items within the Vocabulary and Comprehension subtests of the SDRT between English First Language (EFL) learners and English Additional Language (EAL) learners and Male and Female learners?

1.4. Research Hypotheses

- There is a difference in performance between EFL and EAL learners for both the Vocabulary and Comprehension subtests of the Brown Level of the SDRT.
- There is a difference in performance between male and female learners for both the Vocabulary and Comprehension subtests of the Brown Level of the SDRT.

- There is a difference in performance between EFL and EAL learners on individual items of both the Vocabulary and Comprehension subtests of the Brown Level of the SDRT.
- There is a difference in performance between male and female learners on individual items of both the Vocabulary and Comprehension subtests of the Brown Level of the SDRT.

Chapter 2: Methods

2. 1. Research Design

This is a quantitative study that used archival data collected over the past three years. The research involved a comparison between groups, created by one variable (either language or gender), and a second variable (performance on the SDRT) was measured "to produce a set of scores within each condition" (Gravetter & Forzano, 2003, p.216). It was a non-experimental study because the groups were not created by manipulating an independent variable but rather defined by a pre-existing participant variable (Gravetter & Forzano, 2003). The independent participant variables defining the groups were language, English First Language (EFL) or English Additional Language (EAL), and gender. The dependent variable being measured was the performance on the Vocabulary and Comprehension subtests of the SDRT. The variable was measured only once. The descriptive data did not demonstrate a cause-and-effect explanation for the relationship between the variables but merely described the relationship between the variables.

2.2. Subjects

The purposive sample consisted of 631 Grade Eight learners in three consecutive years. All 631 learners attend a previously termed "Model C" government school in Gauteng and each of the 631 learners have received at least five years of education at an English-medium school. Approximately two-thirds (352 learners), of the subjects spoke English as an Additional Language (EAL) while the other third (279 learners), spoke English as a First Language (EFL). Overall, there was an approximate 50/50 split between male and female learners with 316 males and 315 females. Specifically, there were 154 males and 125 females in the EFL group and 162 males and 190 females in the EAL group (see Table 1). Learners

with severe learning difficulties were excluded from the sample, on the advice of their educators, to ensure that the sample consisted of learners who had a basic reading ability.

Table 1. Descriptive Statistics for the variables of home language and gender for the EFL and EAL groups

GENDER	HOME LANGUAGE		
	EFL	EAL	
MALE	154	162	
FEMALE	125	190	
TOTAL	279	352	***N* = 631**

2.3. Materials

The materials administered in the study were a biographical questionnaire (see Appendix B) and the Vocabulary and Comprehension subtests of the Brown Level of the SDRT. The biographical questionnaire enquired about the subjects' age; their first language and the number of years spent in an English medium school. The Stanford Diagnostic Reading Test is a group-administered test (Karlsen & Gardner, 1986). It is a multiple-choice test and measures "the major components of the reading process" (Karlsen & Gardner, 1986, p. 3), including decoding (phonics and auditory and visual discrimination); vocabulary (word identification and meaning); comprehension (understanding of texts); and rate (pace of reading and ability to skim read).

There are four different levels to the test, the Red Level (Grade One to Three); the Green Level (Grade Four to Five); the Brown Level (Grade Six to Eight); and lastly, the Blue Level (Grade Nine to college or university level) (Karlsen, Madden & Gardner, 1986). The grading of the test reflects the developmental process of reading, which assumes that reading skills and the nature of the reading material become more complex as a child progresses through the school grades (Karlsen et al., 1986). Therefore, the successive levels of the SDRT begin with decoding and progress to increasing levels of vocabulary, comprehension and reading rate (Karlsen et al., 1986). The assumption is that reading comprehension is the most important objective of reading and that vocabulary and decoding are skills required to enhance comprehension (Karlsen & Gardner, 1986). Once comprehension is mastered, the rate at which readers can comprehend should increase (Karlsen & Gardner, 1986).

The SDRT's chief purpose is diagnosing a person's reading strengths and weaknesses (Karlsen et al., 1986). However, it is also used to provide information about reading progress;

identifying learners in need of additional reading support; providing data on the effectiveness of support and intervention programs; and early indication of learner placement (San Diego Unified School District: Standards, Assessment, and Accountability Division, 2003). In addition, the SDRT is different from other reading tests in two ways. Firstly, as a diagnostic instrument the SDRT provides "more detailed coverage of reading skills than does the typical survey test [achievement tests], which covers a broader range of areas in less detail" (Karlsen et al., 1986, p. 5). Secondly, due to the fact that the test "places more emphasis on the low achiever" (Karlsen et al., 1986, p. 5) it consists of more easy questions.

In this study the Brown Level of the SDRT was used because it is applicable to Grade Eight learners. Two subtests were used that are appropriate for the investigation of the SDRT's assessment of reading comprehension. These were the Auditory Vocabulary and Reading Comprehension subtests. The objective of the Auditory Vocabulary is the recognition of "words frequently found in reading materials" (Karlsen & Gardner, 1986, p. 9) that are grade-appropriate. The words are divided into three categories: Reading and Literature; Mathematics and Science; Social Studies and the Arts. At the Brown Level, the purpose of this subtest is to discover if learners understand the meaning of the words that they have learned to decode (Karlsen & Gardner, 1986). Therefore, this subtest assesses listening or auditory vocabulary "to eliminate the decoding aspect" (Karlsen & Gardner, 1986, p. 6).

The Auditory Vocabulary subtest is dictated and provides information about the learners' language competence without requiring them to read (Karlsen et al., 1986). The administrator reads a sentence to the learners, which they must complete from a choice of three words, which are also read, to them. For example, "Courteous means… a *brave* b *polite* c *unusual*". The learner then marks the space for their answer on their answer sheet. This subtest consists of 40 items, which relate to: Reading and Literature (16 items); Mathematics and Sciences (12 items); and Social Studies and the Arts (12 items).

The purpose of the Reading Comprehension subtest is to measure how well the learners can read different types of material (Karlsen & Gardner, 1986). The Reading Comprehension subtest measures two different levels of comprehension: literal and inferential comprehension. It also differentiates between three types of reading, namely, textual; functional, and recreational reading. Literal comprehension refers to explicit meanings in the text while inferential comprehension requires the reader to draw conclusions about implicit meaning. Three different types of text are used: passages from grade-appropriate textbooks

(textual reading); printed material encountered in everyday life (functional reading); and material that is read for pleasure (recreational reading). In the Brown level of this subtest, learners are required to read, independently and silently, a number of short passages and answer the multiple-choice questions that follow. The learner is given 40 minutes within which to complete the subtest. The subtest consists of 30 literal comprehension questions and 30 inferential comprehension questions. The 60 items are further grouped according to the following categories: Textual Reading (20 items); Functional Reading (20 items); and Recreational Reading (20 items).

Tests are scored either manually or electronically. Scores can either be content-referenced or norm-referenced. Therefore, the assessor can choose whether to look at the learner's score on specific sub-tests or compare the score relative to applicable norms. The applicability of the SDRT's norms for all South African learners is questionable. Although the SDRT is reported to be a valid and reliable test in the U.S context (SEDL Reading Resources, 2008), in the South African context, its reliability and validity is unknown. Therefore, this study investigated these psychometric components of the SDRT for the South African context.

The validity of the SDRT as an overall measure of reading ability is commonly recognised (Marsh & Butler, 1984). However, the discriminant validity of the 1966 version of the SDT, in other words, its ability to differentiate between the different components of reading, has been refuted (Marsh & Butler, 1984). Research about subsequent editions' validity is not available. Nevertheless, the SDRT is a reading test that has been used as a valid measure of reading ability in a number of studies (Israel, 2006; Lathy, 2006; Ping-Ha & Chi-Ting, 2000; Pretorius, 2000). In South Africa the test has been used as an instrument in studies centred on the relationship between reading and academic ability (Andor, 2006; Pretorius, 2000).

In this study, the SDRT's construct or internal validity, that is, the extent to which the test produces results that can only be attributed to a single factor, was examined. Validity was determined through item analysis in which item bias may be identified. Item bias refers to those items that may disadvantage different groups (Kanjee, 2001). The present study was interested in whether there are items in the Auditory Vocabulary and Reading Comprehension subtests of the SDRT that might disadvantage South African learners, specifically EFL or EAL learners, as well as male or female learners. The test's validity was also explored by determining if the test was an "equally valid indicator of performance for

both sexes" (Neale, 1989, p. 58). This was achieved by detecting any bias in favour of either sex (Neale, 1989).

2.4. Procedure and Ethical Considerations

In terms of procedure, since the data for the study was archival data, the way in which the data was collected will be described. Prior to data collection, ethical clearance from the Gauteng Department of Education for collecting data in a school and using it for research purposes was applied for and granted. A letter was then sent to the principal of the school describing the purpose and nature of the research (see Appendix A). With the principal's permission, a letter with consent and assent forms was then sent to the parents of the Grade Eight learners. The letter described the purpose of the research study, what it would involve and how the information would be used as well as what feedback would be given to the learners and parents. The letter clearly stipulated that participation was voluntary, that there would be no negative consequences of non-participation and that the confidentiality and anonymity of the learner would be ensured (see Appendix A).

Only those learners from whom parental consent was received and who assented to participate in the research were then sent biographical questionnaires (see Appendix B). Once these questionnaires were collected the subjects were selected on the basis of whether they had received a minimum of five years English medium education.

The sample was then administered the Auditory Vocabulary and Reading Comprehension subtests of the Brown Level of the SDRT. The test was group-administered in classroom settings. The test took approximately 60 minutes to complete. The subjects' tests were then collected, scored and collated. Both a raw score, which indicates how the learner performed on each subtest, and a norm-referenced score was produced for each subject. Only quantitative data was collected.

Confidentiality was maintained throughout the research process and no identifying information appears in the report. There was freedom to withdraw at any time during the research process and there were no negative consequences for this. There were no foreseeable risks for the participants.

Chapter 3: Results

3.1. Data Analysis

The collated data was first, described statistically and then analysed quantitatively. Data gathered from the Auditory Vocabulary and Reading Comprehension subtests of the Brown Level of the SDRT was analysed in reference to the four differentiated groups: Male, Female, EFL, and EAL. Initial assessment of the data revealed that the data was suitable for use with parametric analysis (see Appendix C). Parametric analysis assumes the existence of specific parameters (a summary of values that describes a population (Gravetter & Forzano, 2003) and their estimation (Howell, 1997). This was achieved by assessing whether the scores were normally distributed by using histograms and measures of central tendency. This was followed by a number of statistical techniques designed to address the three research questions.

3.2. Basic Description of Data Set

The descriptive statistics for each independent variable are presented in Table 2.

Table 2: Basic Descriptive Statistics for the two subtests and four groups

Simple Descriptive Statistics

		Vocabulary Subtest (40 items)				
Variable	***N***	***Mean***	***Std. Dev***	***Median***	***Min.***	***Max.***
Female EFL	125	30.30	6.97	31	0	40
Female EAL	190	24.58	6.16	24	0	39
Male EFL	154	29.45	6.27	29.5	0	40
Male EAL	162	25.18	5.84	25	11	39
		Comprehension Subtest (60 items)				
Female EFL	125	45.74	10.42	48	0	60
Female EAL	190	37.66	9.49	37	13	59
Male EFL	154	42.24	10.86	44	0	60
Male EAL	162	34.72	11.15	36	2	59

As indicated in Table 2, the highest mean score for the Vocabulary subtest was for the Female EFL subjects (M = 30.3, SD = 6.97). The lowest mean score was for the Female EAL subjects (M = 24.58, SD = 6.16). Similarly, the Female EFL subjects also achieved the

highest mean score for the Comprehension subtest (*M* = 45.74, *SD* = 10.42), while the Male EAL subjects achieved the lowest mean score (*M* = 34.72, *SD* = 11.15). This suggests that EAL subjects performed below the EFL subjects, as expected.

3.3. Reliability of the Vocabulary and Comprehension Subtests

The first research question sought to determine the internal consistency of the Vocabulary and Comprehension subtests of the SDRT. A reliability coefficient was calculated overall and for each group: EFL learners; EAL learners; Male learners; and Female learners. This was achieved by using data about the number of items in the test; variance of the total test; the proportion of items answered correctly; and the proportion of items answered incorrectly (Wolfaardt, 2001). Internal consistency estimates, indicated by Cronbach Coefficient Alphas were calculated for both subtests of the SDRT. These are presented in Table 3.

Table 3: Reliability estimates for the subtests

	Cronbach's Coefficient Alpha	
Variable	***Raw Variables***	***Standardised Variables***
Overall	.71	.77
Female EFL	.59	.63
Female EAL	.79	.84
Male EFL	.54	.60
Male EAL	.73	.82

Table 3 indicates that, in this study, the SDRT displayed a high level of internal consistency (α = .77) overall (*N* = 631). The SDRT also displayed high levels of internal consistency for female English Additional Language subjects (α = .84) and Male English Additional Language subjects (α = .82). However, the SDRT displayed only moderate levels of internal consistency for female EFL subjects (α = .63) and male EFL subjects (α = .60). While these are still acceptable reliability estimates of internal consistency, the pattern does reflect the fact that the SDRT is aimed at low-achieving readers and, therefore, the consistency in its measurement for EFL subjects is expected to be less than for EAL subjects, whose English Language proficiency is expected to demonstrate greater variance.

3.4. Difference in performance between EFL, EAL, Male and Female subjects

The second research question, which determined the difference in performance on the two subtests of the SDRT between the four groups identified in the sample, namely EFL and EAL learners, and Male and Female learners, was investigated by using a Factorial Analysis of Variance. Factorial or Two-way Analysis of Variance was used because it can determine whether there is a significant difference in mean scores between the EFL and EAL learners and between Male and Female learners and if there is an interaction between the two factors. The Independent Variables that were analysed were Language and Gender, namely: English as a First Language; English as an Additional Language; Male; and Female. The Dependent Variables were scores on the Vocabulary and Comprehensions subtests of the Brown level of the SDRT.

Three hypotheses were tested through the factorial Analysis of Variance. The first hypothesis tested the main effects (the mean difference between the groups' scores) for the language factor, that is, English First Language or English Additional Language. The second hypothesis tested the main effects for the gender factor. And the third hypothesis evaluated the interaction between the factors. A post-hoc analysis was used to determine what was responsible for any differences between the groups.

Results of the Factorial ANOVA with interaction indicated that for the Auditory Vocabulary subtest of the SDRT, there was no significant difference in the performance of male and female subjects ($F(1) = .07$, $p = .8$). However, the difference between the performance of EFL and EAL subjects was highly significant ($F(1) = 97.72$, $p < .0001$). The interaction between the two independent variables (Gender and Home Language) was not significant ($F(1) = 2.06$, $p = .15$) (see Means Plot of Totals: Vocabulary subtest in Appendix D)

For the Reading Comprehension subtest of the SDRT, there was a significant difference in the performance of male and female subjects ($F(1) = 14.65$, $p = .0001$). The difference between the performance of EFL and EAL subjects was also highly significant ($F(1) = 85.81$, $p < .0001$). The interaction between the two independent variables (Gender and Home Language) was not significant ($F(1) = .11$, $p = .74$) (see Means Plot of Totals: Comprehension subtest in Appendix D).

A post-hoc analysis was conducted using the Tukey-Kramer adjustment for multiple comparisons to determine the specific differences between the four groups. The results indicated that there was a significant difference in the performance for the Vocabulary subtest of the SDRT between female EFL and female and male EAL subjects; between female EAL and female and male EFL subjects; between male EFL and female and male EAL subjects; and between male EAL and female and male EFL subjects (see Table 9 in Appendix D).

There was also a significant difference in the performance for the Comprehension subtest of the SDRT between female EFL and female and male EAL, and male EFL subjects; between female EAL and female and male EFL, and male EAL subjects; between male EFL and female EFL, female and male EAL subjects; and between male EAL and female EAL, and male and female EFL subjects (see Table 10 in Appendix D).

3.5. Difference in performance on individual items of the two subtests

Item analysis commonly provides information about the existence of distractors; information about item difficulty; and information about item discrimination power (Murphy & Davidshofer, 1998). In this study, item analysis was primarily used for the purpose of gaining information about item difficultly and to investigate the third research question, which tested the difference in performance on individual items between the groups. The latter was achieved through a Chi-Square table analysis using the *Fisher exact probability test*, which "is based on exact probability rather than reference to a distribution" (Rosenthal & Rosnow, 1991, p. 620), and, thus provided the strictest probabilities.

According to Anastasi (1982, p. 193), the "difficulty of an item is defined in terms of the percentage (or proportion) of persons who answer it correctly". Therefore, the higher the item difficulty index, the easier the item. Item difficulty indices were calculated for all 40 items of the Vocabulary subtest and all 60 items of the Comprehension subtest for each group: Female, Male, EFL, and EAL (see Appendix E). Calculation of item difficulty also provides information about the ability of the item to differentiate between individual differences in level of attainment of what the test is assessing (Anastasi, 1982). For example, an item that no subjects get correct or all subjects get correct is redundant because it provides no information about individual differences (Anastasi, 1982). No items were found to have

an item difficulty index of .0 or 1.0, which means that all the items in both subtests allowed for some differentiation between individuals. Anastasi (1982, p. 193) notes that "the closer the difficulty level of an item approaches .50, the more differentiations the item can make". However, Cohen & Swerdlik (2005) recommend that a test, for which the purpose is to discriminate between individual differences, should contain a spread of items with levels of item difficulty ranging from .40 to .70.

The item analysis also involved observation of ceiling and floor effects that were evident in the scores on different items in the SDRT. A ceiling effect occurs when an item produces a cluster of high scores indicating that a task is too easy (Gravetter & Forzano, 2003). It was unlikely that the sample's scores on the SDRT would produce a ceiling effect. However, it was possible that the data might present a floor effect on certain items of the SDRT. A floor effect is the result of a cluster of low scores indicating that a task is too difficult (Gravetter & Forzano, 2003). Generally, ceiling and floor effects imply "a basic incompatibility between the measurement procedure and the individuals measured" (Gravetter & Forzano, 2003, p. 108). Thus, floor effects in the data provided evidence that certain items of the SDRT might have been unsuitable for use with the sample, particularly given that a feature of the test is that it targets low-achieving readers. The validity of the two subtests was also determined through the observation of ceiling and floor effects which indicated potential item bias.

The third research question required that the individual items of the Vocabulary and Comprehension subtests be analysed for the purpose of determining differences in each group's performance for each item. The results of the Chi-Square table analysis indicated that the EAL subjects performed significantly below the EFL subjects on 63 % of the items in the Vocabulary subtest and 57% of the items in the Comprehension subtest. The females performed significantly below the males on item 17 of the Vocabulary subtest, however they performed significantly above the males on items 1, 8, 13, 45, 46, 47, 52, and 58 of the Comprehension subtest (see Appendix F).

Ceiling effects were identified for several items, particularly for the EFL subjects. However, floor effects were also observed, particularly for the EAL subjects. The following tables demonstrate the item difficulty indices of a number of choice items in the Vocabulary and Comprehension subtests.

Table 4: Item difficulty index: Vocabulary Subtest

Item Difficulty Index (*p*)	**Female EFL**	**Female EAL**	**Male EFL**	**Male EAL**
p9	.57	.37	.58	.29
p10	.63	.41	.57	.34
p15	.37	.22	.41	.30
p22	.37	.28	.34	.29
p26	.50	.37	.65	.38
p30	.48	.31	.46	.36
p36	.48	.22	.36	.16

Table 4 highlights the items in which one or more of the groups demonstrated an item difficulty index under .40. Items with an item difficulty above .70 for one or more of the groups were not included in this table due to the fact that the SDRT aims to identify low-achieving readers and, therefore, the SDRT consists of a number of easy items. It is interesting to note that items 15 and 22 display an item difficulty index below .40 for both EFL and EAL subjects (for example, Female EFL *p22* = .37; Male EFL *p22* = .34). This could imply that these items may hold some bias for South African learners, in general.

Table 5: Item difficulty index: Comprehension Subtest

Item Difficulty Index (*p*)	**Female EFL**	**Female EAL**	**Male EFL**	**Male EAL**
p1	.77	.43	.47	.33
p6	.55	.38	.45	.46
p10	.65	.30	.62	.31
p12	.54	.30	.59	.23
p43	.60	.44	.64	.38
p44	.66	.35	.53	.27
p48	.46	.27	.32	.24
p51	.62	.35	.50	.36
p54	.46	.23	.44	.29
p55	.49	.24	.46	.26
p56	.70	.43	.65	.35
p57	.62	.32	.52	.32

p58	.69	.28	.46	.23
p59	.70	.41	.55	.33

Table 5 highlights the Comprehension items in which one or more of the groups demonstrated an item difficulty index under .40. Items with an item difficulty above .70 for one or more of the groups were not included in this table due to the fact that the SDRT is aimed at identifying low-achieving readers and, therefore, the SDRT consists of a number of easy items. It is significant to note the differences in item difficulty indices between EFL and EAL subjects. The item difficulty indices presented in Table 5 suggest that these items were more difficult for the EAL subjects than for the EFL subjects. However, scanning the content of these items revealed that there might be a number of reasons for the difference in performance.

The statistical analyses that were used to investigate the three research questions were presented in this chapter. The following chapter will discuss the implications of these findings and relate the findings to relevant literature.

Chapter 4: Discussion

In South Africa there is currently scope for extensive research in the area of psychometrics. This study aimed to contribute to this knowledgebase by evaluating the use of the Stanford Diagnostic Reading Test (SDRT) as a measure of Auditory Vocabulary and Reading Comprehension for South African Grade Eight learners. The SDRT is an English Language Proficiency test standardised for the U.S. population. The value of the test, particularly for the South African context, is its emphasis on reading comprehension, with the focus on developing readers. For the majority of South African learners their education, specifically their high school education, is carried out through their second or third language as the medium of instruction. Textbooks used in South African schools are generally either English or Afrikaans. Therefore, English language proficiency is a crucial factor in academic success. In addition, there is a perception in South Africa, that English literacy is associated with empowerment and achievement, both academically and in the workplace (Broom, 2001).

This study investigated the reliability and validity of the Auditory Vocabulary and Reading Comprehension subtests of the SDRT for South African Grade Eight learners; the difference in performance between English First Language (EFL) learners and English Additional Language (EAL) learners on both the subtests; and lastly, individual items in both subtests that demonstrated a significant difference in performance between EFL and EAL learners as well as items that were identified as being notably difficult for all the subjects.

The sample used in the study consisted of 631 Grade Eight learners (made up over three years) from a previously termed 'Model C' high school in Gauteng. All of the subjects had a minimum of five years English medium education. Two-thirds of the subjects were EAL learners and one-third were EFL learners. There were approximately the same number of males and females. The subjects each completed a biographical questionnaire indicating their home language and then completed both the Vocabulary and Comprehension subtests of the SDRT. The data was analysed statistically, using two independent variables, gender and home language.

Overall, the Auditory Vocabulary and Reading Comprehension subtests were found to demonstrate adequate internal consistency ($\alpha = .77$). A reliability coefficient of between 0.7 and 0.9 indicates that the test is reliable for the group for which it was measured (Anastasi & Urbina, 1997). Thus these subtests of the SDRT were found to be generally reliable for use with South African Grade Eight learners. This is consistent with findings of the SDRT's reliability in the U.S.A. (SEDL Reading Resources, 2008). The internal consistency estimates for each of the four subgroups indicated that the two subtests of the SDRT were reliable measures of performance for all the subgroups. However, the estimates suggested that the subtests were slightly more reliable for the EAL learners than the EFL learners. This is explained by the fact that the SDRT is an English Language Proficiency test and, therefore, the majority of EFL learners are expected to perform well on the test whereas, there is expected to be greater variation for EAL learners who may have different levels of English Language Proficiency. Thus, the data demonstrated greater internal consistency for the EAL learners.

To investigate the difference in performance between EFL and EAL learners and male and female learners, a factorial Analysis of Variance (ANOVA) was carried out for the subjects' performance on both the Vocabulary and Comprehension subtest. The results indicated that

there was a significant difference between the performance of EFL and EAL learners on both the Vocabulary ($F(1) = 97.72$, $p < .0001$) and Comprehension ($F(1) = 85.81$, $p < .0001$) subtests of the SDRT. This confirms the first hypothesis, which postulated that EAL learners would perform significantly below EFL learners for both subtests. The same finding was found in previous studies that examined the difference in performance between EFL and EAL learners for these subtests (Andor, 2006; Lathy, 2006). The results are also consistent with international research that indicates that EAL learners whose experience has been one of subtractive bilingualism will perform below their EFL peers due to the consequences of submerging learners in a second language while their first language is still developing (Ben-Zeev, 1984; Cummins, 1991; Miller, 1984).

Specifically, in terms of the lower performance of EAL learners on the Vocabulary subtest, research has shown that bilinguals may have a less developed vocabulary due to sharing their conversation between two language systems, differences in print exposure related to differences in social and economic background, and differences in reading comprehension ability (Ben-Zeev, 1972, 1975; Rosenblum and Pinker, 1983, as cited in Ben-Zeev, 1984; Cunningham & Stanovich, 1991; Reitsma &Verhoeven, 1998, Anderson & Freebody, 1983, as cited in Stanovich et al., 1996; Stanovich & Cunningham, 1992, 1993, as cited in Stanovich et al., 1996). Therefore, EAL learners are confronted with a number of barriers to their vocabulary development, which in turn, influences their ability to comprehend what they are reading.

Subsequently, one of the reasons for the lower performance of the EAL learners for the Comprehension subtest is underdeveloped English vocabulary. However, “the effects of comprehension and vocabulary are reciprocal in that they limit and extend the opportunities for further development in both directions” (Bouwer, 2004, p. 95). This is demonstrated in what is known as “the Matthew effect”, which is named in reference to the parable of the talents in the biblical gospel of Matthew and purports the notion that the rich get richer and the poor poorer. In the context of reading development, “the Matthew effect” is used to refer to the premise that the person who can read well will read more and continue to develop their reading skills. Specifically, the fluent reader expands their vocabulary thus improving their reading skills, which in turn facilitates the further development of their vocabulary and reading ability. Conversely, the poor reader reads less and is less likely to develop their reading skills and extend their vocabulary.

Frequently, for South African EAL learners, learning English is incidental. For many of these learners, English as an additional language is not formally taught and yet they are expected to speak and learn in English as soon as they begin school. Their ability to succeed in learning in English at school is largely dependent on their previous exposure to English. The results of this study suggest that many EAL learners are subject to the repercussions of "the Matthew effect" when they lack early consolidation of English reading skills and consequently struggle to master reading in English. This tends to increase the gap between them and EFL learners.

In terms of gender, there was no significant difference found in the performance of male and female learners for the Vocabulary subtest ($F(1) = .07$, $p = .8$). This suggests that the males and females in this study may have relatively equal print exposure and equal ability in word recognition and identification. Vocabulary scores' correlations with measures of general intelligence also implied that the male and female subjects demonstrated no significant difference in general intelligence (Ben-Zeev, 1984). It is interesting that there was only one item in the Vocabulary subtest, item 17, on which the males performed significantly above the females. The word tested in item 17 is *'elevated'*. There is no obvious reason why the word *'elevated'* was more familiar to the male learners.

The results for the Reading Comprehension subtest indicated that the male learners performed significantly below the female learners ($F(1) = 14.65$, $p = .0001$). This is consistent with research that has shown that males tend to present with reading difficulties more frequently than females (Chiu &McBride-Chang, 2006; Rutter et al., 2004). Furthermore, scanning the items identified as having a significant difference in performance between males and females did not reveal any qualitative information suggesting item bias on the basis of gender. Therefore, it may be assumed that these items indicate some of the difficulties with reading comprehension that males may experience. For example, both the male EFL and EAL learners performed significantly below female learners on items 13 and 47 (see Appendix F), which are inferential comprehension questions and require some knowledge about literacy conventions with reference to choosing a title for a text.

Another factor that could have influenced the performance of learners in the Reading Comprehension subtest is knowledge about techniques to assist in answering comprehension and multiple-choice type questions. Generally, it is possible for a person to answer a literal

comprehension question correctly, without understanding the meaning of all the words in the question. A person is able to do this by finding the same words in the text and looking to see what follows or precedes these words and thus, finding the answer. For example, item 49 of the Reading Comprehension subtest asks, "*In which part of Pennsylvania do the Mennonites live?*" The answer is found in the following sentence taken from the text: "*About 30,000 Amish and Mennonite people live in the "Pennsylvania Dutch" country in south-eastern Pennsylvania.*" The testee is assisted further by the four multiple-choice options that are given: **a** *Northeas*t **b** *South East* **c** *Southwest* **d** *Northwest.* The answer is **b**, because the word 'south-eastern' was found next to the words in the question such as "Pennsylvania" and "Mennonites". In addition, the other choices for the answer were not found in the text. Therefore, it is possible to answer the question correctly without knowing anything about Pennsylvania or knowing the meaning of the word "Mennonites". However, some learners may not know about these techniques and may panic when they see words that they do not recognise and this leads to a poor performance on the test.

An item analysis determining the item difficulty of the items in both the Vocabulary and Comprehension subtests revealed that a number of items in the subtests demonstrated an item difficulty index of less than .40. The EAL learners scored significantly below the EFL learners on 25 of the 40 items in the Vocabulary subtest, ten of which were Reading and Literature words, eight were Mathematics and Science words and seven were Social Studies and Art words. Thus, it seems that the greatest gap in the EAL learners' vocabulary was vocabulary commonly used in Reading and Literature such as *"infant"* (item 16), *"mysterious"* (item 37), *and "an heroic act"*(item 40), which suggests that EAL learners may not read as much English literature as their EFL peers. This is not surprising considering that EAL learners may not read English texts with the same ease as that of EFL learners. In addition, research has shown that parents' beliefs about their ability to help their children read may influence children's reading achievement (Lynch, 2002). It is possible that, given South Africa's political and social past, EAL learners' parents may have lower self-efficacy regarding their ability to support their children's English reading development, thus impacting on the EAL learners' self-perceptions about their reading and subsequently their reading achievement.

It was also interesting to note, in view of detecting items that are biased against all South African learners, that items 15 and 22 in the Vocabulary subtest displayed an item difficulty

index below .40 for both EFL and EAL subjects. However, scanning the content of these items suggests that this was not the case. The words tested in these items appeared to hold no sign of bias, as they were words that South African learners learning in English may be expected to know (see Table 6 below).

Table 6: Vocabulary Items with an item difficulty index less than .40 for all groups

Vocabulary Item	***Item difficulty Index (p)***			
	Female EFL	***Female EAL***	***Male EFL***	***Male EAL***
15: ecology	.37	.22	.41	.30
22: adhesive	.37	.28	.34	.29

Table 6 indicates that only 37% of female EFL, 41% of male EFL, 22% of female EAL and 30% of male EAL learners understood the meaning of the word "ecology". Similarly, only 37% of female EFL, 34% of male EFL, 28% of female EAL and 29% of male EAL learners understood the meaning of the word "adhesive". This suggests that the learners in this study might not have been exposed to literature in which these words would commonly be found. It could also relate to the content and timing of the school curriculum. Perhaps the subject of ecology is only introduced in later grades.

Item analysis of the subjects' performance on the Comprehension subtest revealed that there was a significant difference between the performance of EFL and EAL subjects for 34 items out of a total of 60 items. Scanning the content of these items indicated that many of the items for which there was a significant difference between EFL and EAL learners, consisted of content that may be described as inappropriate for the South African context. For example, item 6 asks, "How much would a senior citizen pay in the afternoon?" The four choices are: **e** 75c **f** $2.00 **g** $2.50 **h** $3.50. The use of the dollar symbol is inappropriate for use in South Africa. However, on closer inspection of the item, this may not be the reason that EAL learners did not get this item correct. The item requires the testee to refer to the provided passage which is a cinema advertisement specifying the prices of tickets for different groups of people at different times. The differences in cost of tickets for the three groups of people (children, adults, and senior citizens) are differentiated in terms of "Before 5:00 P.M. and After 5:00 P.M). On first glance one might see the word "After" and assume this was the column specifying the price of tickets in the afternoon. However, this is not the case, since

the afternoon is before 5 P.M. This example illustrates the difficulty of identifying item bias because without a more detailed distractor analysis it is impossible to be completely certain as to the reason why a particular group of individuals found this item difficult.

Nevertheless, there are several other items on which the EAL learners performed significantly below the EFL learners, in which inappropriate content may be identified. Item 11 refers to a "mitt" and item 12 refers to "baseball", both of these terms may be unfamiliar to the majority of South African learners since baseball is not a sport that is popularly played in South Africa. However, there must be caution in making such assumptions because as was highlighted in the literature review, South Africans are at different stages of westernisation, urbanisation and literacy (Shuttleworth-Jordan, 1996). Therefore, many of the EAL learners in this study may have had exposure to baseball and be familiar with the baseball kit such as the mitt. Still, it is important that assessment measures be as relevant as possible to the population for which they are being used.

Another example of a text used in the Comprehension subtest that contains unsuitable content for South African learners is a passage about 'snowshoeing'. This passage has few contextual similarities with the context of the majority of South African learners. The passage refers to 'hunting boots', 'pine woods' and 'white hills'. All of which are objects that are unfamiliar to the majority of South African learners. Another passage is about the Amish community in Pennsylvania and refers to the "buggy". This passage is highly removed from the context of South African learners. For item 48, which was related to this passage and included the word "buggies", only 46% of female EFL learners, 32% of male EFL learners, 27% of female EAL learners and 24% of male EAL learners correctly answered this item.

Hence, there is no doubt that inappropriate content in some of the passages and questions in the Comprehension subtest of the SDRT account for the lower performance of the EAL learners, whose level of westernisation may not be the same as that of their EFL counterparts. However, there is also evidence from the literature that suggests that the difference in performance between EFL and EAL learners could also be attributed to the fact that most of the EAL learners have experienced a form of subtractive bi- or multilingualism. Consequently, their acquisition of English has been to the detriment of both their first language as well as to their progress in English due to the fact that they have been submerged in English from an early age, whilst their first language was still developing.

Evidence of underdeveloped English language proficiency may be indicated in the lower performance of EAL learners on inferential comprehension items in the subtest. For example, the EAL learners performed significantly below the EFL learners on the last seven questions, all of which were inferential questions, of the Comprehension subtest. The passage to which these questions were related was a functional passage about viewing a solar eclipse. There was no inappropriate content in this passage that could account for the lower performance of the EAL learners. However, the questions did require a more extensive vocabulary due to the fact that they were inferential questions. This could also account for EAL learners' lower performance on other items of the subtest.

Based on the preceding discussion, it seems that the validity of the Vocabulary and Comprehension subtests of the SDRT for South African learners remains intact despite there being a number of items that are deemed to contain inappropriate content. Since the SDRT is an English language proficiency test, it is expected that EAL learners will perform below their EFL counterparts. Furthermore, the difference in performance between males and females on the Comprehension subtest seems to reflect research that has shown that reading difficulties are more predominant in males. In addition, the Vocabulary and Comprehension subtests were found to be reliable overall and for each of the four groups, namely Female EFL, Female EAL, Male EFL, Male EAL. However, given the character of the test, the EFL learners' performance demonstrated less variance and thus a lower internal consistency estimate.

4.1. Limitations of the study

The limitations of this study centre largely on the existence of possible confounding variables that were not excluded from this study. It has been argued that the lower performance of the EAL learners on the Vocabulary and Comprehension subtests of the SDRT is accounted for by the fact that they are second or additional language English speakers. This is certainly a reasonable argument to make given the function of the SDRT as an English language proficiency test. However, there are a number of other factors, which may have influenced the performance of both the EFL and EAL learners. Variables such as socio-economic background, previous exposure to English and different reading materials, quality of previous schooling, practice with comprehension and multiple-choice type tests, and general cognitive and academic ability would significantly affect a person's success in a test of this nature.

This means that it is impossible to determine whether difference in performance on the two subtests of the SDRT was solely a result of language or whether it was a product of the aforementioned factors.

Another limitation of this study was that there was no opportunity to conduct a distractor analysis. A distractor analysis "involves examining the frequency with which each incorrect response is chosen" (Murphy & Davidshofer, 1998, p. 195) by the group tested. In this way, distractors (incorrect responses) that are most commonly used may be identified. Item difficulty is largely determined by the credibility of the distractor (Murphy & Davidshofer, 1998). Therefore, it stands to reason that if a testee has no or little conceptualisation of the content being tested "any distractor may be equally plausible" (Murphy & Davidshofer, 1998, p. 203). Consequently, by identifying distractors, content that is inappropriate for the South African context could have been highlighted and, thus, a more detailed guideline for identifying which items could be adapted through further research could have been provided. Lastly, the internal validity of the study could have been potentially compromised by personal variables of the subjects such as boredom, fatigue and the tendency to guess answers.

4.2. Theoretical and Practical Implications of the study

The research did make a number of positive and valuable theoretical and practical contributions. Firstly, the paucity of research on the SDRT, as the primary focus of a study, means that this study represents a starting point for research in this direction. In addition, the study forms part of a burgeoning body of research evaluating the use of standardised psychological and educational tests in the South African context. Secondly, the results of the study emphasise the chasm between EFL and EAL learners' English language proficiency, which is a crucial factor in academic success in South Africa. Regardless of whether this difference is attributed to home language or other socio-economic factors, the results indicate that action is required to support EAL learners in the English classroom.

Furthermore, this study highlighted the negative effects of a subtractive form of bilingualism, which is the case for many South African EAL learners. And therefore, further advances the cause for a whole language policy to be implemented in South African schools, including the right to first language instruction. There is no doubt that this is idealistic, however, the emphasis should be on ensuring that South African learners begin to experience a more

additive form of bilingualism. Broom (2001) recommends that South African language-in-education policy should differentiate between short-term and long-term goals. Short-term goals should include establishing EAL learners', particularly those living in urban areas, first language literacy and supporting their transition to English medium education, while at the same time affirming the constructive use of other languages in the classroom. The long-term goal for language-in-education policies is additive bi/multilingualism for both EAL learners and EFL learners, which requires enhancing "the perceived status of African languages both in education and in society as a whole" (Broom, 2001, p. 212).

The research also drew attention to the prevalence of reading difficulties, particularly difficulties with reading comprehension, amongst the male population. This could also be accounted for in terms of the amount of print exposure based on gender stereotypes as well as level of interest in the content of reading passages, enjoyment of reading; personal and parents' self-perceptions about reading; and the school and peer culture related to reading (Bray & Barron, 2003; Chiu & McBride-Chang, 2006; Lynch, 2002; Pomplun & Omar, 2001).

Finally, this study identified a number of items and passages in the Comprehension subtest of the Brown level of the SDRT that were found to contain potentially inappropriate content for the South African context. Although, there is no evidence that this content constitutes as item bias, it is recommended that content such as the dollar symbol, a passage about snow shoeing, and a passage about the Amish community in Pennsylvania be adapted or substituted with more contextually relevant texts to promote the suitability of the SDRT for the South African context.

4.3. Further Research

Future research could endeavour to replicate these results having matched the subjects in terms of socio-economic background, quality of previous schooling, and general cognitive and academic ability. It would be interesting to investigate the correlation between academic achievement and performance on the SDRT for South African high school learners. Longitudinal studies are necessary to evaluate the success of current South African language-in-education polices being implemented (Broom, 2001). Further studies investigating the male prevalence of reading difficulties would be interesting, particularly within the South

African context, taking into consideration the effects of gender and race stereotyping that impact on self-esteem and motivation.

Research that utilises a distractor analysis to identify which incorrect responses are chosen by both EFL and EAL learners in the Vocabulary and Comprehension subtests of the SDRT would facilitate further detection of items that may demonstrate content bias, thus providing a quantitative basis for adapting these items. Lastly, this research forms part of the foundation for future research into standardised psychological and educational tests used in South Africa for the current South African population. Studies that examine test adaptation in South Africa are particularly pertinent, with the emphasis on making tests appropriate for the South African context while at the same time maintaining the test's reliability.

4.4. Conclusion

There is no doubt that the Comprehension subtest of the SDRT contains content that could and should be adapted to promote the use of the SDRT in South Africa. However, the SDRT cannot necessarily be described as being biased against a certain group in South Africa, especially because it is an English language proficiency test and therefore, should accurately differentiate between individuals' English reading abilities. This study has found the Vocabulary and Comprehension subtests of the Brown level of the SDRT to do this effectively for a sample of Grade Eight learners. However, further research is required to test whether the results of this study can be generalised to other groups of learners in South Africa and to explore the possibilities of adapting or making substitutions of certain items in the Comprehension subtest.

Shuttleworth-Jordan's (1996) caution against the hasty abandonment of westernised assessment measures is merited with significant credence. It is too easy to conclude that a test is unsuitable for use with the South African population due to the fact that one group of individuals performs below another group, and therefore, the test is culturally irrelevant to that group or that the group is heavily disadvantaged due to their culture or language. Such a statement involves several assumptions about a particular group of people. Indeed, there is danger of discriminating against individuals in the endeavour to make a test fair to all individuals. A truly bias free or culture fair test cannot exist. Difference may only be made in the degree to which a test is able to measure what it is intended to measure for different people, regardless of race, culture, and language.

References

Anastasi, A. (1982). *Psychological Testing.* 5th Edition. New York: MacMillan Publishing Company.

Anastasi, A. & Urbina, S. (1997). *Psychological Testing.* 7th Edition. New Jersey: Prentice-Hall.

Andor, K.E. (2006). *First and Second Language English Students in Tertiary Education: Reading and academic performance.* Unpublished Masters Research Report. Johannesburg: University of the Witwatersrand.

Beech, J.R. & Singleton, C. (1997). The Psychological Assessment of Reading: Theoretical issues and professional solutions. In Beech, J.R. and Singleton, C. (Eds.*). The Psychological Assessment of Reading*. New York: Routledge.

Ben-Zeev, S. (1984). Bilingualism and Cognitive Development. In Miller, N. (Ed.). *Bilingualism and Language Disability: assessment and remediation.* London: Croom Helm, pp. 55 - 80

Bouwer, C. (2004). Reading and Writing. In Eloff, I. & Ebersöhn, L. (Eds.). *Keys to Educational Psychology,* pp. 83 – 118.

Bowyer-Crane, C. & Snowling, M.J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology,* 75, pp. 189-201.

Bray, G.B. & Barron, S. (2003). Assessing Reading Comprehension: The effects of text-based interest, gender and ability. *Educational Assessment,* 9, (3 &4), pp. 107 - 128.

Broom, Y.M. (2001). *Becoming Literate in a Multilingual Society: Learning to read in South African schools.* Unpublished Doctoral Thesis. Johannesburg: University of the Witwatersrand.

Chiu, M.M. & McBride-Chang, C. (2006). Gender, Context, and Reading: A Comparison of Students in 43 Countries. *Scientific Studies of Reading,* 10, (4), pp. 331 –362.

Cohen, R.J. & Swerdlik, M. (2005). *Psychological Testing and Assessment: An Introduction to tests and measurement.* 6th Ed. USA: McGraw Hill Higher Education.

Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. *Journal of Education,* 163, (1), pp 16-29.

Cummins, J. (1991). Language Development and Academic Learning. In Malavé, L.M. & Duquette, G. (Eds.). *Language, Culture and Cognition: a collection of studies in first and second language acquisition.* Clevedon: Multilingual Matters, pp. 161-175

Cummins, J. & Mulcahy, R. (1978). Orientation to Language in Ukranian-English Bilingual Children. *Child Development,* 49, (4), pp. 1239 – 1242.

Department of Education. (1997). *Quality Education for All: Overcoming Barriers To Learning and Development: Report of the National Commission on Special Needs in Education and Training (NCSNET), National Committee and Educational Support Services (NCESS).* Pretoria: Department of Education.

Department of Education. (2001). Education White Paper 6. *Special Needs Education: Building an Inclusive Education and Training System.* Pretoria: Department of Education.

Duquette, G. (1991). Cultural Processing and Minority Language Children with Needs and Special Needs. In Malavé, L.M. & Duquette, G. (Eds.). *Language, Culture and Cognition: a collection of studies in first and second language acquisition.* Clevedon: Multilingual Matters, pp. 120 - 136

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

Foxcroft, C. & Roodt, G. (2001). An overview of assessment: definition and scope. In Foxcroft, C. & Roodt, G. (Eds.). *An Introduction to Psychological Assessment in the South African context.* Cape Town: Oxford University Press, pp. 3 - 10.

Foxcroft, C., Roodt, G. & Abrahams, F. (2001). 'Psychological Assessment: a brief retrospective overview' in Foxcroft, C. and Roodt, G. (Eds.). *An Introduction to Psychological Assessment in the South African context*. Cape Town: Oxford University Press, pp. 11 – 33.

Gough, P.B., Hoover, W.A., & Peterson, C.L. (1996). Some observations on a simple view of reading. In Cornoldi, C. and Oakhill, J. (Eds.). *Reading Comprehension Difficulties: Processes and Intervention.* New Jersey: Lawrence Erlbaum Associates, pp. 1 – 13.

Gravetter, F.J. & Forzano, L.B. (2003). *Research Methods for the Behavioural Sciences.* California: Wadsworth/ Thomson Learning.

Gregory, R.J. (2007). *Psychological Testing: History, Principles and Applications*. 5th Edition. Boston: Pearson Education.

Hambleton, R.K. (1996, April). *Guidelines for Adapting Educational and Psychological Tests.* Paper presented at the Annual Meeting of the National Council of Measurement in Education, New York. Sponsored by the National Center for Education Statistics, Washington, DC.

Hemp, F. (1989). Neuropsychological impairment in children following head injury. Unpublished Doctoral dissertation. Cape Town: University of Cape Town.

Howell, D.C. (1997). *Statistical Methods for Psychology*. 4th edition. London: Wadsworth Publishing Company.

Israel, N. (2006). *Raven's Advanced Progressive Matrices within a South African context. Unpublished Masters Research Report*. Johannesburg: University of the Witwatersrand

Kanjee, A. (2001). 'Cross-cultural test adaptation and translation' in Foxcroft, C. and Roodt, G. (Eds). *An Introduction to Psychological Assessment in the South African context.* Cape Town: Oxford University Press, pp. 86 – 102.

Karlsen, B. & Gardner, E.F. (1986). *Stanford Diagnostic Reading Test: Manual for interpreting. Brown Level.* (3rd Ed.). New York: The Psychological Corporation.

Karlsen, B., Madden, R. & Gardner, E.F. (1986). *Stanford Diagnostic Reading Test: Directions for Administering. Brown Level.* (3rd Ed.). New York: The Psychological Corporation.

Kessler, C. (1984). Language Acquisition in Bilingual Children. In Miller, K. (Ed.). *Bilingualism and Language Disability: assessment and remediation.* London: Croom Helm, pp. 26 - 54

Kriegler, S.M. & Skuy, M. (1996). 'Perspectives on psychological assessment in South African schools' in Engelbrecht, P., Kriegler, S,M. & Booysen, M.I. (Eds.). *Perspectives on Learning Difficulties: International concerns and South African realities.* Pretoria, South Africa: Van Schaik.

Lathy, H.L. (2006). *An assessment of reading in First Language (L1) and Second Language (L2) learners who experience barriers to learning.* Unpublished Masters Research Report. Johannesburg: University of the Witwatersrand.

Legaretta, D. (1979). The Effects of Program Model on Language Acquisition by Spanish Speaking Children. *TESOL Quarterly,* 13, (4), pp. 521 – 534.

Lerner, J. (1993). *Learning Disabilities: Theories, Diagnosis and Teaching Strategies.* 6th edition. Boston: Houghton Mifflin Company.

Leyba, C.F. (1978). *Longitudinal Study Title VII Bilingual Program Santa Fe Public Schools, Santa Fe, New Mexico.* Los Angeles: California State University, National Dissemination and Assessment Centre.

Lynch, J. (2002). Parents' self-efficacy beliefs, parents' gender, children's reader self-perceptions, reading achievement and gender. *Journal of Research in Reading,* 25, (1), pp. 54 - 67.

Machet, M.P. (1991). The Effect of Cultural Background on Reader Response and Memory of Text Content. *South African Journal of Higher Education,* 5, (1).

Marsh, H.W. & Butler, S. (1984). Evaluating Reading Diagnostic Tests: An application of Confirmatory Factor Analysis to Multitrait-Multimethod Data. *Applied Psychological Measurement,* 8, (3), pp. 307 – 320.

Martin, D. (1997). Towards a New Multilingual Language Policy in Education in South Africa: different approaches to meet different needs. *Educational Review,* 49, 2, pp. 129 – 139.

Miller, N. (1984). Language Problems and Bilingual Children. In Miller, N. (Ed.). *Bilingualism and Language Disability: assessment and remediation.* London: Croom Helm, pp. 81 - 103

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

Neale, M.D. (1989). Neale Analysis of Reading Ability – Revised. British Edition. *Manual.* Oxford: NFER – Nelson Publishing Company.

Oakhill, J. & Cain, K. (1998). Problems in text comprehension: current perspectives and recent research. In Reitsma, P. & Verhoeven, L. (Eds.). *Problems and Interventions in Literacy Development.* Dordrecht, the Netherlands: Kluwer Academic Publishers, pp. 177-192

Owen, K. (1998). *The Role of Psychological Tests in Education in South Africa: issues, controversies and benefits.* Pretoria, RSA: Human Sciences Research Council.

Ping-Ha, C. & Chi-Ting, C. (2000). Evaluating Sustained Silent Reading in Reading Classes. *The Internet TESL Journal,* 11.

Pluddemann, P. (1999). Multilingualism and education in South Africa: One year on. *International Journal of Educational Research,* 31, pp. 327-340.

Pomplun, M. & Omar, M.H. (2001). Do reading passages about war provide factorially invariant scores for men and women? *Applied Measurement in Education,* 14, (2), pp. 171 – 189.

Pretorius, E.J. (2000). "What they can't read will hurt them": reading and academic achievement. *Innovation,* 21. Editorial Committee, c/o University Library, University of Natal, pp. 33-41.

Reitsma, P. & Verhoeven, L. (1998). Problems in Literacy Acquisition and Interventions. In Reitsma, P. & Verhoeven, L. (Eds.). *Problems and Interventions in Literacy Development.* Dordrecht, the Netherlands: Kluwer Academic Publishers, pp. 3-12

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

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

Rosier, P. & Farella, M. (1976). Bilingual Education at Rock Point: Some Early Results. *TESOL Quarterly,* 10, (4), pp. 379 – 388.

Rutter, M., Caspi, A., Fergusson, D.M., Horwood, L.J., Goodman, R., Maughan, B., Moffitt, T.E., Carroll, J. (2004) Gender differences in reading difficulties: Findings from four epidemiology studies. *Journal of the American Medical Association,* 291, pp. 2007-2012.

San Diego Unified School District: Standards, Assessment, and Accountability Division. Stanford Diagnostic Reading Test (SDRT) Summaries: 2003-04. Accessed 6 February 2008.**http://studata.sandi.net/research/sdrt/index.asp**

SEDL Reading Resources. *The Stanford Diagnostic Reading Test.* 4th edition. Accessed 15 March 2008. http://www.sedl.org/cgi-bin/mysql/rad.cgi?searchid=228

Shuttleworth-Jordan, A.B. (1996). On not reinventing the wheel: A clinical perspective on culturally relevant test usage in South Africa. *South African Journal of Psychology*, 26, 2, pp. 96 – 102

Skuy, M., Schutte, E., Fridjhon, P. & O'Carroll, S. (2001). Suitability of Published Neuropsychological Test Norms for Urban African Secondary School Students in South Africa. *Personality and Individual Differences, 3*, pp. 1413-1425.

Stanovich, K.E., West, R.F., Cunningham, A.E., Cipielewski, J., & Siddiqui, S. (1996). The role of inadequate print exposure as a determinant of reading comprehension problems. In Cornoldi, C. and Oakhill, J. (Eds.). *Reading Comprehension Difficulties: Processes and Intervention.* New Jersey: Lawrence Erlbaum Associates, pp. 15 – 32.

Valdés, G. & Figueroá, R.A. (1994). *Bilingualism and testing: A special case of bias.* Norwood, NJ: Ablex.

Weber, S. & Tardif, C. (1991). Culture and Meaning in French Immersion Kindergarten. In Malavé, L.M. & Duquette, G. (Eds.). *Language, Culture and Cognition: a collection of studies in first and second language acquisition.* Clevedon: Multilingual Matters, pp. 93 - 109

Wolfaardt, J.B. (2001). 'Basic Concepts' in Foxcroft, C. and Roodt, G. (Eds). *An Introduction to Psychological Assessment in the South African context.* Cape Town: Oxford University Press, pp. 34 – 68.

APPENDIX A

A.1. Information Sheet

A.2. Consent Form

A.3. Assent Form

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 [redacted] 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 [redacted] High School. 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, [redacted] or via e-mail 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

Ms [redacted]

Consent Form

I, the parent/ guardian of _____ consent to my son/daughter taking part in the reading research to be conducted at High School under the supervision of 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 _____

Assent Form

I _____ agree to take part in the reading research to be conducted at _____ High School under the supervision of _____ 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 B

B.1. 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 C

C.1. Normality of the Data

Normality of the Data

The assessment of whether the data was suitable for parametric analysis was achieved by evaluating the normality of the distribution of the data, through histograms, measures of central tendency and Kolmogorov-Smirnov Tests for Normality.

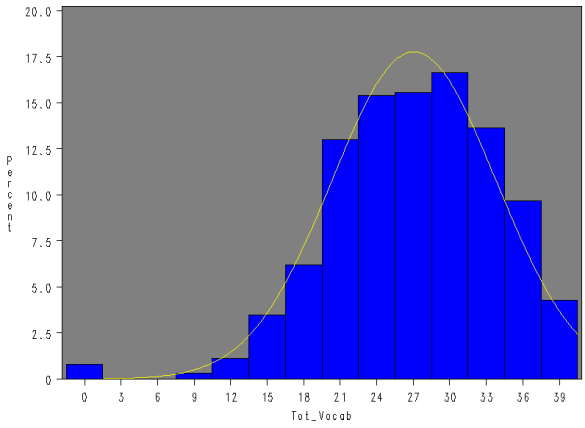

Figure 1: Histogram: SDRT Vocabulary

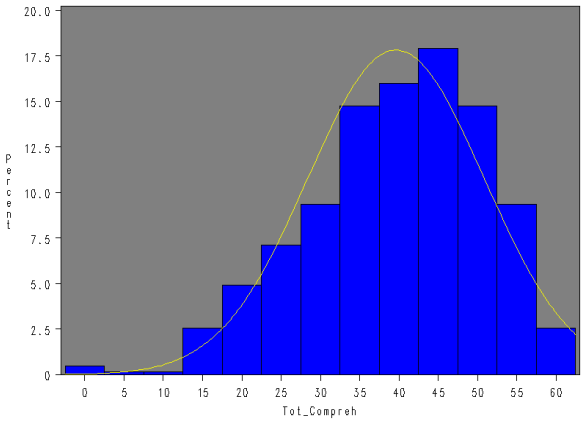

Figure 2: Histogram: SDRT Comprehension

Examination of the two histograms for the two subtests of the SDRT revealed that both the Auditory Vocabulary subtest and the Reading Comprehension subtest data were slightly negatively skewed (refer to Figure 1 and Figure 2), indicating that the majority of subjects performed relatively well on the test. This was expected, given the nature of the test and the fact that the test aims to identify low-achieving readers. These visual representations of the score distribution were confirmed by an assessment of the measures of central tendency (to locate the centre of the distribution of scores) for the two subtests, with specific emphasis on the distance between the mean and the median.

Table 7: Basic Descriptive Statistics for the two subtests

Simple Descriptive Statistics

Variable	*N*	*Mean*	*Std. Dev*	*Median*	*Mode*	*Minimum*	*Maximum*
Vocab.	631	27.06	6.74030	27	29	0	40
Compr.	631	39.63	11.19295	40	39	0	60

As observed in Table 7, this assessment tended to confirm that the Auditory Vocabulary subtest and the Reading Comprehension subtest were sufficiently normally distributed to allow for parametric analysis (Vocabulary: Mean = 27.06; Median = 27; Comprehension: Mean = 39.63; Median = 40).

Table 8: Kolmogorov-Smirnov Tests for Normality for the Subtests

Kolmogorov-Smirnov Goodness-of-fit test for Normal Distribution		
	Statistic (D)	***p – Value***
Vocab.	0.05558079	$p < .010$
Compr.	0.06565018	$p = .009$

According to the results of the Kolmogorav-Smirnov Tests, as indicated in Table 8, neither the Vocabulary nor the Comprehension subtest could be considered normally distributed (Vocabulary: $p < .010$; Comprehension: $p = .009$). However, Howell (1997) notes that these tests are particularly sensitive to discrepancies, and, furthermore, no extraordinarily high levels of significance were found.

Therefore, although it seems that there may be a slightly negatively skewed distribution of scores for both subtests the measures of central tendency found that the data revealed adequate normal distribution for parametric analysis.

APPENDIX D

D.1. Means Plots of Totals: Vocabulary Subtest

D.2. Means Plots of Totals: Comprehension Subtest

D.3. Post-hoc Analysis

D.1. Means Plots of Totals: Vocabulary Subtest

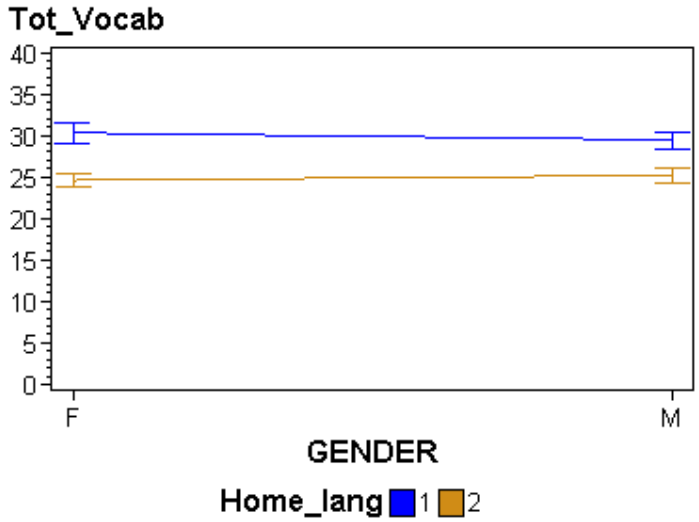

Figure 3: Means plot of Totals for Vocabulary subtest by Gender and Home Language

D.2. Means Plots of Totals: Comprehension subtest

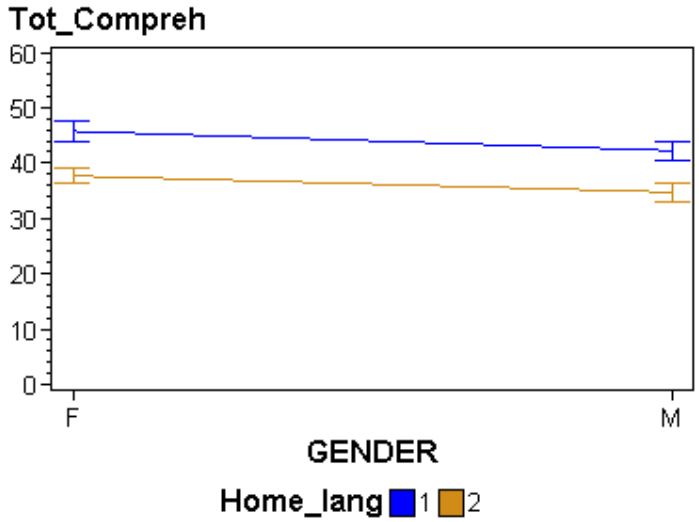

Figure 4: Means plot of Totals for Comprehension subtest by Gender and Home Language

Figure 3 and Figure 4 graphically present the mean plots for the Vocabulary and Comprehension subtests, respectively for the variables of home language and gender. It is clear that in both subtests the EAL subjects performed below the EFL subjects. Figure 3 illustrates there was no significant difference in the performance of male and female subjects for the Vocabulary subtest. However, Figure 4 indicates that for the comprehension subtest the male subjects performed below the female subjects.

D.3. Post-hoc Analysis

Table 9: Post hoc analysis: Tukey-Kramer – Vocabulary Subtest

	Significance (p)			
	Female EFL	Female EAL	Male EFL	Male EAL
Female EFL	--	< .0001	.67	< .0001
Female EAL	< .0001	--	< .0001	.81
Male EFL	.67	< .0001	--	< .0001
Male EAL	< .0001	.81	< .0001	--

Table 10: Post hoc analysis: Tukey-Kramer – Comprehension Subtest

	Significance (p)			
	Female EFL	Female EAL	Male EFL	Male EAL
Female EFL	--	< .0001	.03*	< .0001
Female EAL	<.0001	--	.0004	.04*
Male EFL	.03*	.0004	--	<.0001
Male EAL	<.0001	.04*	<.0001	--

* $p < .05$

APPENDIX E

E.1. Item Difficulty Index: Vocabulary subtest

E.2. Line: Percentage Correct Items: Vocabulary subtest

E.3. Item Difficulty Index: Comprehension subtest

E.4. Line Graph: Percentage Correct Items: Comprehension subtest

E.1. Item Difficulty Index: Vocabulary subtest

Table11: Item difficulty index: Vocabulary Subtest

Item Difficulty Index (*p1 – p40*)	**Female EFL**	**Female EAL**	**Male EFL**	**Male EAL**
p1	.72	.456	.623	.519
p2	.872	.758	.922	.809
p3	.888	.80	.825	.784
p4	.904	.826	.909	.864
p5	.88	.70	.805	.753
p6	.864	.821	.909	.852
p7	.592	.537	.558	.593
p8	.84	.711	.773	.642
p9	.568	.367	.584	.290
p10	.632	.405	.565	.346
p11	.912	.847	.89	.827
p12	.888	.721	.851	.778
p13	.912	.863	.948	.92
p14	.944	.958	.955	.932
p15	.368	.216	.409	.296
p16	.76	.563	.669	.50
p17	.632	.421	.792	.580
p18	.776	.537	.792	.580
p19	.736	.595	.695	.617
p20	.832	.611	.812	.642
p21	.688	.474	.636	.469
p22	.368	.284	.344	.290
p23	.912	.916	.870	.864
p24	.704	.495	.701	.519
p25	.80	.579	.818	.556
p26	.504	.368	.649	.377
p27	.824	.690	.766	.753
p28	.848	.758	.786	.784
p29	.824	.658	.812	.642
p30	.48	.305	.461	.358

p31	.928	.884	.961	.944
p32	.912	.737	.916	.803
p33	.888	.821	.916	.852
p34	.808	.742	.708	.772
p35	.944	.795	.877	.778
p36	.48	.221	.357	.161
p37	.824	.632	.766	.630
p38	.60	.505	.520	.420
p39	.792	.637	.708	.654
p40	.656	.442	.591	.432

E.2. Line Chart: Percentage Correct Items: Vocabulary subtest

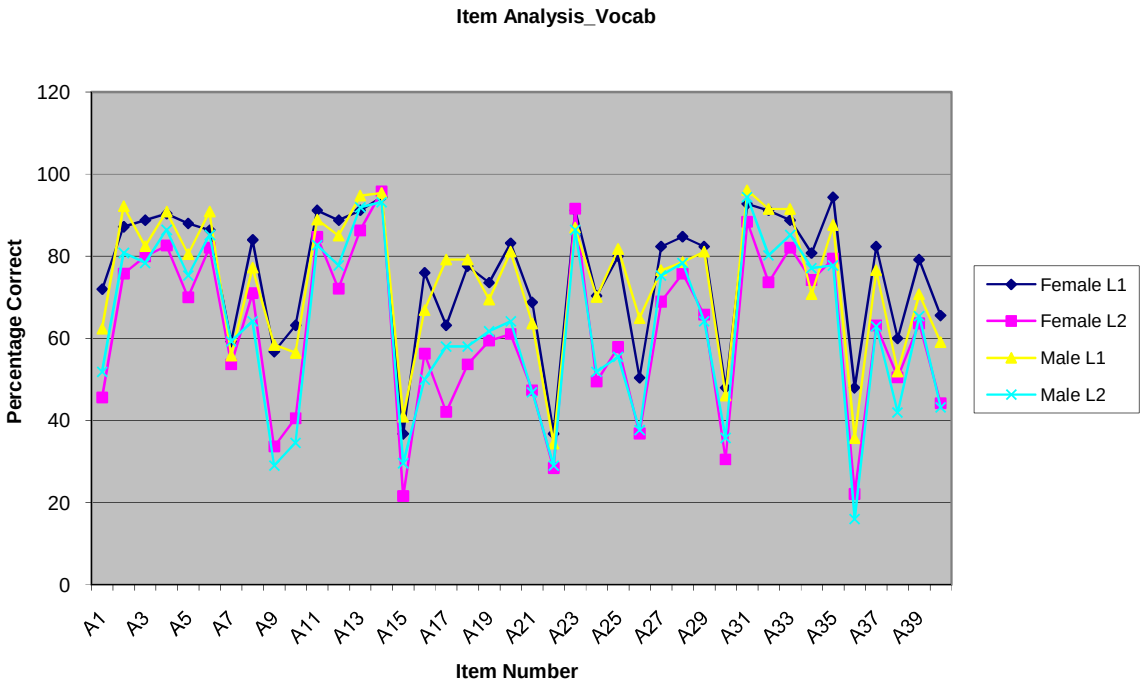

Figure 5: Plot Graph: Percentage Correct for each item for each group: Vocabulary subtest

E.3. Item Difficulty Index: Comprehension subtest

Table 12: Item difficulty index: Comprehension Subtest

Item Difficulty Index (*p1 – p60*)	**Female EFL**	**Female EAL**	**Male EFL**	**Male EAL**
p1	.768	.426	.474	.333
p2	.824	.758	.838	.728
p3	.816	.737	.857	.778
p4	.568	.49	.565	.531
p5	.648	.579	.675	.568
p6	.552	.384	.448	.463
p7	.912	.70	.831	.716
p8	.728	.590	.643	.414
p9	.968	.926	.948	.858
p10	.648	.30	.617	.309
p11	.856	.605	.779	.648
p12	.544	.295	.591	.228
p13	.792	.711	.675	.562
p14	.968	.932	.948	.914
p15	.656	.637	.682	.549
p16	.712	.505	.662	.537
p17	.88	.879	.877	.877
p18	.96	.932	.955	.858
p19	.768	.737	.753	.673
p20	.648	.616	.701	.531
p21	.944	.868	.961	.809
p22	.864	.853	.825	.889
p23	.592	.433	.565	.457
p24	.84	.784	.773	.698
p25	.776	.742	.773	.654
p26	.912	.89	.909	.858
p27	.632	.611	.688	.506
p28	.634	.574	.623	.494
p29	.936	.942	.896	.901

p30	.856	.832	.831	.728
p31	.76	.56	.675	.537
p32	.88	.847	.89	.778
p33	.864	.884	.877	.858
p34	.896	.868	.89	.864
p35	.872	.721	.831	.716
p36	.92	.895	.89	.796
p37	.784	.737	.662	.648
p38	.864	.826	.844	.809
p39	.912	.674	.805	.636
p40	.704	.642	.636	.605
p41	.84	.684	.786	.580
p42	.80	.611	.695	.580
p43	.60	.437	.643	.377
p44	.656	.353	.533	.272
p45	.80	.595	.623	.488
p46	.736	.584	.597	.451
p47	.856	.80	.675	.636
p48	.456	.274	.318	.241
p49	.864	.726	.773	.648
p50	.832	.695	.714	.611
p51	.624	.353	.50	.315
p52	.808	.558	.604	.50
p53	.736	.516	.636	.432
p54	.456	.226	.435	.290
p55	.488	.242	.455	.259
p56	.704	.426	.649	.352
p57	.616	.321	.52	.321
p58	.688	.284	.455	.228
p59	.704	.411	.546	.333
p60	.832	.616	.721	.494

E.4. Line Chart: Percentage Correct Items: Comprehension subtest

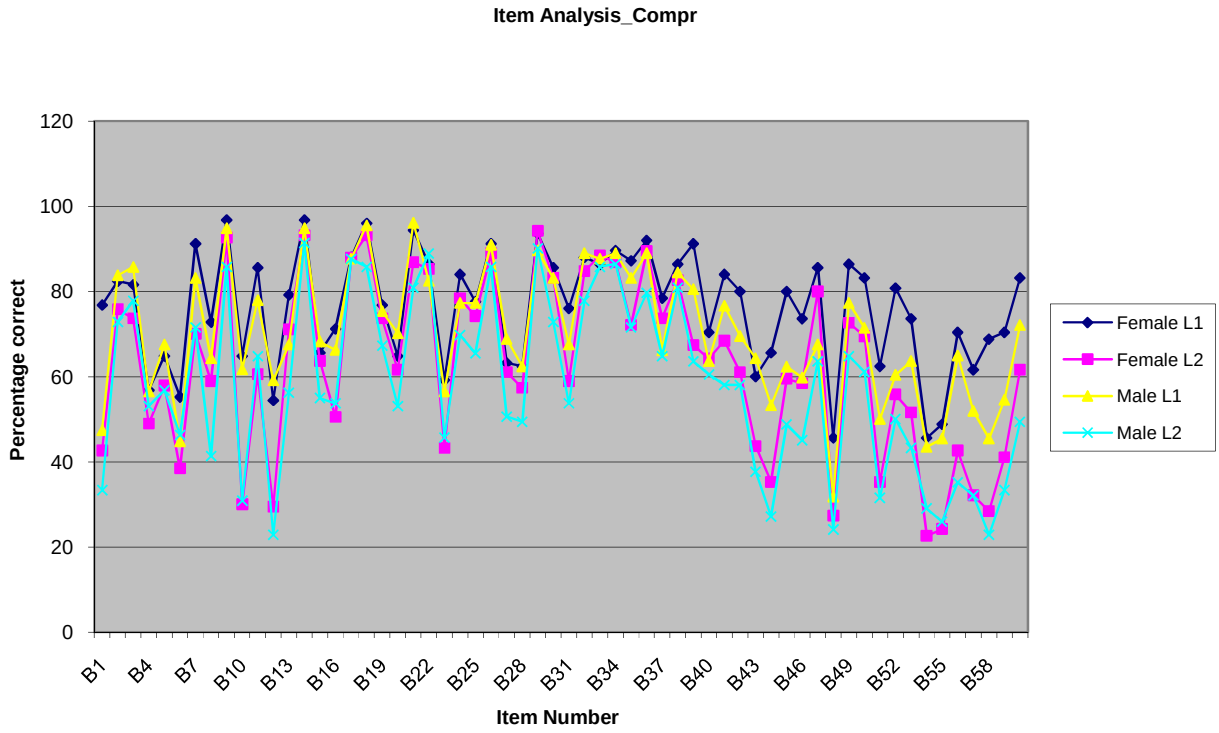

Figure 6: Plot Graph: Percentage Correct for each item for each group: Comprehension subtest

APPENDIX F

F.1. Fisher Exact P-values

F.1. Fisher Exact P-values

Table 13: Chi Square Test: Significant differences in performance on individual items

Fisher Exact P-value			
Vocabulary Subtest			
Item	***HOME LANGUAGE***	***Item***	***GENDER***
1	0	17	.00001
2	.00006		
5	.00075		
8	.00055		
9	0		
10	0		
12	.00019		
15	.00028		
16	.00001		
17	0		
18	0		
19	.00542		
20	0		
21	0		
24	0		
25	0		
26	0		
29	0		
30	.00041		
32	0		
35	.00004		
36	0		
37	.00001		
39	.00725		
40	.00001		
Comprehension Subtest			
Item	***HOME LANGUAGE***	***Item***	***GENDER***
1	0	1	.00007
2	.0087	8	.00278
7	0	13	.00086
8	.00002	45	.00186
9	.00412	46	.00214
10	0	47	0
11	0	52	.00721
12	0	58	.00712
16	.00003		
18	.00608		
21	0		
23	.00133		
31	.00013		
35	.00011		

39	0
41	.00001
42	.00014
43	0
44	0
45	.00006
46	.00063
48	.00135
49	.00046
50	.00277
51	0
52	.00003
53	0
54	0
55	0
56	0
57	0
58	0
59	0
60	0

APPENDIX G

G.1. Letter of Approval from Faculty

Faculty of Humanities - Postgraduate

Private Bag 3, Wits 2050, South Africa • **Tel:** +27 11 717 4002/8 • **Fax:** +27 11 717 4037 • **E-mail:** Julie.Poyser@wits.ac.za

Student Number: 329154

11 August 2008

Dear Ms Catto

APPROVAL OF PROPOSAL FOR THE DEGREE OF MASTER OF EDUCATION IN EDUCATIONAL PSYCHOLOGY BY COURSEWORK AND RESEARCH REPORT

I am pleased to be able to advise you that the readers of the Graduate Studies Committee have approved your proposal entitled *"Evaluating the Stanford diagnostic reading test as a measure of reading comprehension and vocabulary for South African grade eight learners"* and you have now been admitted to full candidature. I confirm that Dr Y Broom has been appointed you supervisor in the Department of Psychology.

The research report is normally submitted to the Faculty Office by 15 February. All students are required to RE-REGISTER at the beginning of each year. Your registration form for the year 2008 will be posted to you at the end of the year.

You are required to submit 2 bound copies and 2 unbound copies (loose pages) of your research report to the Faculty Office. The 2 bound copies go to the examiners and are retained by them and the 2 corrected ~~corrected~~ unbound copies are eventually sent to Archives and to the Library.

Please note that should you miss the deadline of 15 February you will be required to submit an application for extension of time and register for the research report extension. Any candidate who misses the deadline of 15 February will be charged fees for the research report extension.

I should be glad if you keep us informed of any changes of address during the year.

Note: All MA and PhD candidates who intend graduating shortly must meet your ETD requirements at least 6 weeks prior to graduation.

Yours sincerely

Julie Poyser
Postgraduate Division
Faculty of Humanities
Private Bag X3
Wits, 2050
Tel: +27 11 717 4008
Fax:+27 11 717 4037