
THE STANFORD DIAGNOSTIC READING TEST: ASSESSING ENGLISH READING COMPREHENSION AND VOCABULARY IN SOUTH AFRICA

Yvonne Broom
University of the Witwatersrand
Deborah Diane Jewson

This study evaluated the Stanford Diagnostic Reading Test (SDRT) as a measure of vocabulary and reading comprehension for South African Grade Eight learners. Two subtests, Auditory Vocabulary and Reading Comprehension, of the Brown level of the SDRT were administered to three consecutive years of Grade Eight learners attending a former Model C school in Gauteng. All of the 631 subjects had received at least five years of English medium education. The sample included 279 English First Language (EFL) learners and 352 English Additional Language (EAL) learners with 316 males and 315 females. The EAL learners performed significantly below the EFL learners on both the Vocabulary and Reading Comprehension subtests of the SDRT. There was a significant difference in performance between male and female learners on the Comprehension subtest but not on the Vocabulary subtest. Both subtests demonstrated internal consistency although their validity may be compromised by items that demonstrated bias against EAL learners. The suitability and utility of this test in the multilingual South African context is discussed.

INTRODUCTION

The Stanford Diagnostic Reading Test (SDRT) is an English Language Proficiency test that was first published in 1966 (Karlsen & Gardner, 1986). The Fourth Edition, published in 1996, is appropriate for use with English First Language speakers from Grade One to Grade Twelve. The SDRT uses diagnostic criteria to provide detailed information about different reading skills and it 'places more emphasis on the low achiever' (Karlsen & Gardner, 1986:3), which allows for a more accurate assessment of learners with varying English language abilities. In South Africa this is particularly important as the majority of learners are not first language English speakers but are being educated in English (Barry, 2001; Heugh, 2009). The SDRT is one of the few formal reading tests that measure different levels of reading comprehension (Karlsen & Gardner, 1986, 1994 & 1995). In South Africa, most learners have mastered decoding skills by Grade Eight but many struggle with reading comprehension, especially the English Additional Language (EAL) learners (Pretorius, 2010). Difficulty in reading comprehension is a barrier to learning that precludes effective use of textbooks particularly when 'learning to read' gives way to 'reading to learn' (Pretorius, 2010). Therefore, a test such as the SDRT that identifies a learner's strengths and areas for growth in terms of different levels of reading comprehension is valuable.

The SDRT, as an American-developed test, has been standardised for the US population only. Foxcroft and Roodt (2005) highlight the relationship between assessment measures and their context. 'Assessment measures are usually developed in a certain context (society or 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 re-norming the measure' (Foxcroft & Roodt, 2005:5). The suitability of the SDRT in the South African context, therefore, requires consideration.

The reliability and validity of a test can be compromised when it is used in a context other than the one in which it was developed, particularly when it is administered to people from a different culture to those for whom the test was devised. Cross-cultural testing involves attempting to eliminate from a test those features that may disadvantage people whose beliefs, knowledge, perceptions and experiences are different to those for whom the test was designed (Anastasi & Urbina, 1997). For example, the value attached to the speed of completing a test is a factor that varies across culture and, therefore, cross-cultural tests attempt to limit the influence of time on performance. Test content is frequently culture-specific so the content of cross-cultural tests needs to be considered carefully to ensure that people from different cultures have the experiential background to understand the content (Anastasi & Urbina, 1997).

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 cultural or socio-economic groups (an option that is neither practical nor economically viable and could be described as a type of discrimination) (Owen, 1998). There are some psychologists who advocate that assessment measures either be modified or newly created for specific cultural groups (Helms, 1992 & Davidson, 1995). In response to Davidson (1995), Dyck (1996) 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 is potential for discrimination even under the pretext of making a measure more culturally appropriate.

Shuttleworth-Jordan (1996:96) further 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, 1996:96). She argues that in South Africa people are at different stages of urbanisation, Westernisation and literacy. She notes that differences in test performance are frequently attributed to cultural differences when it is, in fact, differences in levels of education that have led to the discrepancy (Skuy, Schutte, Fridjhon & O'Carroll, 2001; Shuttleworth-Edwards, Kemp, Rust, Muirhead, Hartman & Radloff, 2004). However, culture does 'dictate what is and what is not relevant, and provides models for ways of thinking, acting and feeling' (Shuttleworth-Jordan, 1996:97-98) which influence test performance.

Shuttleworth-Jordan (1995) conducted a comparative research study that compared the performance of African first language and English first language students of equivalent

university education on a range of non-verbal and verbal tests (SAWAIS: 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. In explaining the poorer performance of the African first language students, Shuttleworth-Jordan argued 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. Further studies have supported the quality of education as having greater impact than language on performance of South African students on educational and neuropsychological tests (Skuy, Schutte, Fridjhon & O'Carroll, 2001; Shuttleworth-Edwards et al., 2004). Shuttleworth-Jordan (1996) 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.

Finding ways to assess every individual fairly is not an easy task. Standardised tests have significant value in terms of being well researched, valid and reliable measures of well-defined constructs. It is a challenge to find ways to adapt tests so that they become more appropriate for the context but retain their reliability and validity.

This study investigated 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 Grade Eight learners. A number of studies, both internationally and locally, have used the SDRT (Carter, 2012; Lipka & Siegel, 2012; Ping-Ha & Chi-Ting, 2000; Pretorius, 2000; Winnett, 2008), but the appropriateness of using the SDRT with first and second language English speakers in South Africa has not been investigated. The performance of English First Language (EFL) and English Additional Language (EAL) learners on the Vocabulary and Comprehension subtests of the SDRT was explored. The study focused on whether there are items in the Vocabulary and Comprehension subtests of the SDRT that might disadvantage either EFL or EAL South African learners. The test's validity was further explored by determining if the test was an 'equally valid indicator of performance for both sexes' (Neale, 1989:58). In addition, 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, 2005).

The following research questions were investigated:

1. Are the Vocabulary and Comprehension subtests of the SDRT reliable measures of performance for EFL and EAL learners in South Africa?
2. Did the performance of the EFL and EAL learners on the Vocabulary and Comprehension subtests of the SDRT differ?
3. Was there a difference between the performance of the Male and Female learners on the Vocabulary and Comprehension subtests of the SDRT?
4. Did performance of EFL and EAL Male and Female learners differ on individual items of the Vocabulary and Comprehension subtests of the SDRT?

In the discussion of the results some underlying reasons for the differences or similarities in the performances of these groups, and the suitability of the SDRT for use with South African learners are considered.

Method

This was a longitudinal quantitative study with data collected over three years. The independent participant variables defining the groups were language, English First Language (EFL) or English Additional Language (EAL) and gender. The dependent variable was performance on the Vocabulary and Comprehension subtests of the SDRT. 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.

The purposive sample consisted of 631 Grade Eight learners in three consecutive years. All 631 learners attended an English medium, government school in Johannesburg, South Africa. Each of the learners had received at least five years of education at an English medium school prior to the study. The learners had all attended the same primary schools in the area and were from similar socio-economic backgrounds. Approximately a third of the learners (279 learners) spoke English as a First Language (EFL) while the remainder (352 learners) spoke English as an Additional Language (EAL) and their home language was one of the official African languages.¹ There were 316 male and 315 female learners, with 154 males and 125 females in the EFL group and 162 males and 190 females in the EAL group (see Table 1).

Table 1 Home language and gender of the Grade 8 participants

Gender	English First Language	English Additional Language	Total
Male	154	162	316
Female	125	190	315
Total	279	352	631

Materials

The measures administered in the study were a biographical questionnaire and the Vocabulary and Comprehension subtests of the Brown Level of the SDRT (Karlsen & Gardner, 1994). The biographical questionnaire enquired about the learners' age, first language and the number of years spent in an English medium school. The SDRT is a group-administered multiple-choice test that measures 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). The grading of the test reflects the developmental process of reading that assumes that reading skills and the nature of the reading material become more complex as a child progresses through the school grades (Karlsen & Gardner, 1994). Therefore, the successive levels of the SDRT begin with decoding, and progress to increasing levels of vocabulary, comprehension and reading rate (Karlsen, Madden & Gardner, 1986). The assumption is that reading comprehension is the objective of reading and that vocabulary and decoding are skills required to enhance comprehension (Cain, Oakhill & Lemmon, 2004; Geva & Farnia, 2012; Joshi, 2005; Karlsen & Gardner, 1994; Ricketts, Nation & Bishop, 2007). Once comprehension is mastered, the rate at which readers can comprehend should increase (Karlsen & Gardner, 1994).

The SDRT's major purpose is diagnosing 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 test], which covers a broader range of areas in less detail' (Karlsen et al., 1986:5). Secondly, the test contains some easy items because it 'places more emphasis on the low achiever' (Karlsen et al., 1986:5).

The Brown Level of the SDRT which is recommended for learners in their eighth year of formal education was used in this study. The objective of the Auditory Vocabulary subtest is the recognition of 'words frequently found in reading materials' (Karlsen & Gardner, 1986:9). The words of the 40 items are divided into three categories: Reading and Literature, Mathematics and Science, and Social Studies and the Arts. This 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 by choosing one of three words that are also read to them. For example, 'Courteous means... (a) *brave* (b) *polite* (c) *unusual*'. The learner then marks their answer on an answer sheet. The Reading Comprehension subtest determines how well the learners can read different types of material (Karlsen & Gardner, 1986). It requires both literal and inferential comprehension and differentiates between three types of reading: textual, functional and recreational. 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 and 30 inferential comprehension questions.

Procedure

Permission was obtained to collect data in the school from the Gauteng Department of Education and from the school principal and teachers. The learners were informed about the research and those who consented to participate were assessed at the end of their Grade Eight year. The assessment was conducted by researchers and class teachers in the classrooms who follow the administration instructions for the SDRT (Karlsen & Gardner, 1994).

Results

The performance of the Grade Eight participants on the SDRT are summarised in the descriptive statistics for each independent variable and shown in Table 2.

Table 2 Performance of the Grade Eight learners on the Vocabulary and Comprehension subtests of the SDRT

	Descriptive Statistics		
	N	Mean	SD
	Vocabulary Subtest (40 items)		
Female EFL	125	30.03	6.97
Female EAL	190	24.58	6.16
Male EFL	154	29.45	6.27
Male EAL	162	25.18	5.84
	Comprehension Subtest (60 items)		
Female EFL	125	45.74	10.42
Female EAL	190	37.66	9.49
Male EFL	154	42.24	10.86
Male EAL	162	34.72	11.15

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 indicates that EAL subjects performed more poorly than the EFL subjects.

The first research question concerned determination of 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 using data about the number of items in the test, the variance of the total test, the proportion of items answered correctly, and the proportion of items answered incorrectly (Wolfaardt & Roodt, 2005). Internal consistency estimates, indicated by Cronbach Coefficient Alphas, were calculated for both subtests of the SDRT.

Table 3 Reliability estimates for the SDRT subtests

Variable	Cronbach's Coefficient Alpha	
	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 ($\alpha = .77$) overall ($N = 631$), as well as for female EAL learners ($\alpha = .84$) and Male EAL learners ($\alpha = .82$). However, only moderate levels of internal consistency were found for female EFL subjects ($\alpha = .63$) and male EFL subjects ($\alpha = .60$). While these are 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 learners would be expected to be less than for EAL learners whose English Language proficiency demonstrated greater variance.

The second and third research questions considered the difference in performance on the two subtests of the SDRT between the four groups of EFL and EAL learners, and Male and Female learners. The data was found to be suitable for parametric analysis through histograms, measures of central tendency and Kolmogorov-Smirnov Tests for Normality. Results of the Factorial ANOVA with interaction indicated that for the Auditory Vocabulary subtest of the SDRT there was no significant difference between the performance of male and female subjects ($F(1) = .07, p = .8$). However, the difference between 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 ($F(1) = 2.06, p = .15$).

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

A post-hoc analysis was conducted using the Tukey-Kramer adjustment for multiple comparisons to determine the specific differences between the four groups (see Tables 4 and 5). The results confirm the significant difference in the performance on the Vocabulary and Comprehension subtests of the SDRT between the four groups of participants. The differences were mainly between the different language groups rather than the different genders within the language groups.

Table 4 Post-hoc analysis: Tukey-Kramer – Vocabulary Subtest

	Significance (<i>p</i>)			
	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	--

There was a significant difference in the performance on the Comprehension subtest between female EFL and female and male EAL, and male EFL participants; between female EAL and female and male EFL, and male EAL learners; between male EFL and female EFL, female and male EAL learners; and between male EAL and female EAL, and male and female EFL learners.

Table 5 Post hoc analysis: Tukey-Kramer – Comprehension Subtest

	Significance (<i>p</i>)			
	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	--

The fourth research question concerned the validity of the individual items of the Vocabulary and Comprehension subtests of the SDRT. Item analysis commonly provides information about distractors, item difficulty, and item discrimination power (Murphy & Davidshofer, 2004). In this study, item analysis was used to gain information about item difficulty and to investigate the difference in performance on individual items between the groups. This was

achieved through a Chi-Square table analysis using the Fisher exact probability test, which provided the strictest probabilities.

According to Anastasi and Urbina (1997:193), 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. Calculation of item difficulty also informs the ability of the item to differentiate between individual differences in level of attainment of what the test is assessing (Anastasi & Urbina, 1997). For example, an item that was not answered correctly by any participants, or one that all participants get correct, is redundant because it provides no information about individual differences. 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 and Urbina (1997:193) note that ‘the closer the difficulty level of an item approaches .50, the more differentiations the item can make.’ However, Cohen and Swerdlik (2005) recommend a spread of items with levels of difficulty ranging from 0.40 to 0.70.

The validity of the two subtests was also determined through the observation of ceiling and floor effects, which indicated potential item bias. A ceiling effect occurs when an item produces a cluster of high scores indicating that the item is too easy (Gravetter & Forzano, 2003). A floor effect indicates that an item is too difficult. Generally, ceiling and floor effects imply ‘a basic incompatibility between the measurement procedure and the individuals measured’ (Gravetter & Forzano, 2003:108). Ceiling effects were identified for several items, particularly for the EFL learners. However, floor effects were also observed, particularly for the EAL learners. The following tables demonstrate the item difficulty indices of a number of choice items in the Vocabulary and Comprehension subtests.

The results of the Chi-Square table analysis indicate 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 were significantly worse than males on item 17 of the Vocabulary subtest, but significantly better than males on items 1, 8, 13, 45, 46, 47, 52 and 58 of the Comprehension subtest.

Table 6 Item difficulty index: Vocabulary Subtest

Item number	Female EFL	Female EAL	Male EFL	Male EAL
9	.57	.37	.58	.29
10	.63	.41	.57	.34
15	.37	.22	.41	.30
22	.37	.28	.34	.29
26	.50	.37	.65	.38
30	.48	.31	.46	.36
36	.48	.22	.36	.16

Table 6 indicates the items in which one or more of the groups of participants demonstrated an item difficulty index less than .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 notable that items 15 and 22 display an item difficulty index below .40 for both EFL and EAL

learners (for example, Female EFL 22 = .37; Male EFL 22 = .34). This implies that these items may hold some bias for South African learners in general.

Table 7 Item difficulty index: Comprehension Subtest

Item number	Female EFL	Female EAL	Male EFL	Male EAL
1	.77	.43	.47	.33
6	.55	.38	.45	.46
10	.65	.30	.62	.31
12	.54	.30	.59	.23
43	.60	.44	.64	.38
44	.66	.35	.53	.27
48	.46	.27	.32	.24
51	.62	.35	.50	.36
54	.46	.23	.44	.29
55	.49	.24	.46	.26
56	.70	.43	.65	.35
57	.62	.32	.52	.32
58	.69	.28	.46	.23
59	.70	.41	.55	.33

Table 7 indicates the Comprehension items in which one or more of the groups demonstrated an item difficulty index under 0.40. The differences in item difficulty indices between EFL and EAL learners are notable and indicate which items were more difficult for EAL than for EFL learners.

Discussion

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. 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 performance of English First Language (EFL) learners and English Additional Language (EAL) learners on both the subtests was determined and an analysis of individual items in both subtests was conducted. The results indicated a significant difference in performance between EFL and EAL learners. The extent to which these differences could be explained by the difficulty of individual items was investigated and items that were difficult for all the subjects were identified.

Overall, the Auditory Vocabulary and Reading Comprehension subtests were found to demonstrate adequate internal consistency ($\alpha = .77$) indicating that they are generally reliable for use with South African Grade Eight learners. This is consistent with findings of the SDRT's reliability in the USA (SEDL Reading Resources, 2008).

To investigate the difference in performance between EFL and EAL learners and male and female learners, a factorial Analysis of Variance (ANOVA) was conducted on the Vocabulary and Comprehension subtests. 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. This is consistent with international research indicating that EAL learners who have experienced subtractive bilingualism perform worse than their EFL peers (Ben-Zeev, 1984; Bournot-Trites & Tellowitz, 2002; Cummins, 1991; Low &

Siegel, 2005; Menken & Kleyn, 2010; Miller, 1984). In turn, this consistency with international results suggests that the SDRT is a sensitive and suitable measure of vocabulary and comprehension in both EFL and EAL learners in South Africa. There was no significant difference between 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. Only item 17 in the Vocabulary subtest was answered significantly better by males than 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 of 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 language and reading difficulties more frequently than females (Chiu & McBride-Chang, 2006; Rutter, Caspi, Fergusson, Horwood, Goodman, Maughan, Moffitt & Carroll, 2004). Inspection of the items that differed significantly between males and females did not reveal any suggestion of 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, 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 all the learners in the Reading Comprehension subtest is knowledge about 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, just by finding the same words in the text and looking to see what follows or precedes these words. 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) Northeast (b) South-east (c) Southwest (d) Northwest**. The answer is **b**, because the word *'south-eastern'* was found in the sentence while the other choices were not found in the text. Therefore, it is possible to answer the question correctly without knowing anything about Pennsylvania or Mennonites. However, some learners may not know this technique and may panic when they see unfamiliar words and this could lead to poor performance.

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 0.40. 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 the Grade Eight 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, the parents of EAL learners may be less able and willing to support their children's English reading development.

Access to reading materials might also impact on the EAL learners' exposure to, and proficiency with, English literature. Lukhele (2009) investigated the relationship between access to leisure reading materials and second language reading ability in her study exploring the relationships between reading attitudes, reading ability and academic performance among teacher trainees in Swaziland. She found a significant relationship between access to reading resources and reading ability so 'students who had better access to reading resources tended to perform better on the reading test' (Lukhele, 2009:137). She highlighted that in Swazi society reading is primarily associated with education not leisure and that 'reading books for leisure is a most uncommon practice' (Lukhele, 2009:6). Exposure to English literature is another factor that may cause South African EAL learners to demonstrate poorer English reading comprehension than EFL learners in the same Grade.

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 unfamiliar in South Africa but this may not be the reason that EAL learners did not get this item correct. The item requires comprehension of 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 by 'Before 5:00 P.M.' and 'After 5:00 P.M.' At first glance one might see the word 'After' and assume this was 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 certain about the reason why some 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 South African learners since baseball is not a sport that is popularly played in South Africa. However, such assumptions should be made with caution as South Africans are at different stages of Westernisation, urbanisation and literacy (Shuttleworth-Jordan, 1996), so some EAL learners may have had exposure to baseball. Still, it is important that assessment measures be as relevant as possible to the population for which they are being used.

A further example of a text used in the Comprehension subtest that contains unsuitable content for South African learners is a passage about 'snowshoeing'. The passage refers to 'hunting boots', 'pine woods' and 'white hills', all of which would be unfamiliar to the majority of South African learners. Another passage about the Amish community in Pennsylvania is far removed from the context of South African learner and Item 48, which related to this passage, was only answered correctly by 46% of female EFL learners, 32% of male EFL learners, 27% of female EAL learners and 24% of male EAL learners. It seems possible that inappropriate content in some of the passages and questions in the Comprehension subtest might impact on the poorer performance of the EAL learners. This

matter could be investigated further by conducting a distractor analysis that ‘involves examining the frequency with which each incorrect response is chosen’ (Murphy & Davidshofer, 1998:195) by the group tested. Distractors (incorrect responses) that are most commonly used may be identified, and if a testee has little or no conceptualisation of the content being tested, ‘any distractor may be equally plausible’ (Murphy & Davidshofer, 1998: 203). Thus, by identifying distractors, content that is inappropriate for the South African context might have been better identified.

Evidence of poor 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 relating to a passage about viewing a solar eclipse – a topic not unfamiliar to the learners. The sensitivity of the SDRT to the differences between the EAL and EFL learners suggests that it is a useful and appropriate instrument for determining the vocabulary and comprehension abilities of South African learners. It appears that the validity and reliability of the Vocabulary and Comprehension subtests for South African learners remain intact despite there being a number of items that contain inappropriate content. 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 (Chiu & McBride-Chang, 2006; Rutter et al., 2004).

Limitations of the study

The limitations of this study centre largely on possible confounding variables that were not investigated. 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, particularly in the context of South African education (Pretorius & Cummins, 2010). Attempts were made to minimise the influence of many of these factors by selecting participants from the same school with similar socio-economic backgrounds and educational experiences. In addition, the participants all had at least five years of English medium education, which would mean that they would all have had exposure to comprehension and multiple-choice type tests. However, the generalisability of these results to other contexts in South Africa should be considered with caution due to the wide diversity of educational and language situations.

Theoretical and Practical Implications

The research makes a number of positive theoretical and practical contributions to the assessment of reading comprehension in South Africa. The study contributes to a growing body of research evaluating the use of standardised psychological and educational tests in multicultural contexts and suggests that the observed differences between performance by EAL and EFL are not artefacts of the assessment process and instrument used, but reflect real differences requiring further research to investigate their roots and implications for education in South Africa. The research highlights the prevalence of reading comprehension difficulties

amongst learners, and in particular amongst male learners which could be accounted for by the amount of print exposure based on gender stereotypes as well as level of interest in reading, 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; Poplin & Omar, 2001).

This study identified a number of items and passages in the Comprehension subtest of the Brown level of the SDRT that contain potentially inappropriate content for the South African context; this may also be the case in other countries, and it is recommended that content such as the dollar symbol, the passages about snowshoeing and the Amish community be adapted or substituted with more contextually relevant texts to promote the suitability of the SDRT in multicultural contexts. Awareness that inclusion of these items in the South African context may lead to poorer performance than expected from the standardised norms developed in other countries is essential and highlights the importance of a relevant comparison group.

In conclusion, this study has found the Vocabulary and Comprehension subtests of the Brown level of the SDRT to effectively differentiate between individuals' English reading abilities for a sample of Grade Eight learners. Shuttleworth-Jordan's (1996) caution against the hasty abandonment of Westernised assessment measures is supported by this research. It is too easy to conclude that a test is unsuitable for use with a particular population because one group of individuals performs below another group, and, therefore, that the test is culturally irrelevant and may disadvantage the group. Indeed, a truly bias-free or culture-fair test for a diverse population such as that in South Africa may not be attainable, and a more appropriate focus should be on the extent to which a test is able to measure what it is intended to measure for different people, regardless of race, culture, and language.

¹ There are 11 officially recognised spoken languages in South Africa: Afrikaans, English, IsiNdebele, IsiXhosa, IsiZulu, Northern Sotho, Sesotho, Setswana, SiSwati, Tshivenda, Xitsonga.

REFERENCES

- ANASTASI, A & S URBANI. 1997. *Psychological Testing*. (7th ed). New Jersey: Prentice-Hall.
- BARRY, D 2001. Language equity and assessment in South African education. *Journal for Language Teaching*, 36(1&2):105-117.
- BEN-ZEEV, S. 1984. Bilingualism and Cognitive Development. In Miller, N (Ed), *Bilingualism and Language Disability: assessment and remediation*. London: Croom Helm. 55-80.
- BOURNOT-TRITES, M & U TELLOWITZ. 2002. Report of Current Research on the Effects of Second Language Learning on First Language Literacy Skills. A Report Commissioned by the Atlantic Provinces Educational Foundation. Halifax, Nova Scotia: The Printing House.
- BRAY, GB & S BARRON. 2003. Assessing Reading Comprehension: The effects of text-based interest, gender and ability. *Educational Assessment*, 9(3):107-128.
- CAIN, K, J OAKHILL & K LEMMON. 2004. Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, 96:671-681.

- CARTER, K. 2012. The effects of a reading intervention on first and second language English medium learners. Unpublished Masters Research Report. Johannesburg: University of the Witwatersrand.
- CHIU, MM & C McBRIDE-CHANG. 2006. Gender, Context, and Reading: A Comparison of Students in 43 Countries. *Scientific Studies of Reading*, 10(4):331-362.
- CHOW, SSR & C CHOU. 2000. Evaluating Sustained Silent Reading in Reading Classes. *The Internet TESL Journal*, 6(11). Available <http://iteslj.org/Articles/Chow-SSR.html>.
- COHEN, RJ & M SWERDLIK. 2005. *Psychological Testing and Assessment: An Introduction to tests and measurement*, (6th ed.). USA: McGraw Hill Higher Education.
- CUMMINS, J. 1991. Language Development and Academic Learning. In Malave, LM & G Duquette (Eds), *Language, Culture and Cognition: a collection of studies in first and second language acquisition*. Clevedon: Multilingual Matters. 161-175.
- DAVIDSON, G. 1995. Cognitive assessment of indigenous Australians: towards a multiaxial model. *Australian Psychologist*, 30(1):30-34.
- DYCK, M.J. 1996. Cognitive assessment in a multicultural society: comment on Davidson (1995). *Australian Psychologist*, 31(1): 66-69.
- FOXCROFT, C & G ROODT. 2005. An overview of assessment: definition and scope. In Foxcroft, C & G Roodt (Eds), *An Introduction to Psychological Assessment in the South African context* (2nd ed). Cape Town: Oxford University Press.3-7.
- GEVA, E & F FARNIA. 2012. Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading & Writing*, 25(8): 819-1845.
- GRAVETTER, FJ & LB FORZANO. 2003. *Research Methods for the Behavioural Sciences*. California: Wadsworth/Thomson Learning.
- HELMS, JE. 1992. Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, 47(9):1083-1101.
- HEUGH, K. 2009. Contesting the monolingual practices of a bilingual to multilingual policy. *English teaching Practice and Critique*, 8(2):96-113.
- JOSHI, RM. 2005. Vocabulary: A critical component of comprehension. *Reading & Writing Quarterly*, 21:209-219.
- KANJEE, A. 2005. Cross-cultural test adaptation and translation. In C Foxcroft & G Roodt (Eds), *An Introduction to Psychological Assessment in the South African context* (2nd ed). Cape Town: Oxford University Press. 57-66.
- KARLSEN, B & EF GARDNER. 1986. *Stanford Diagnostic Reading Test: Manual for interpreting*. Brown Level. (3rd ed). New York: The Psychological Corporation.
- KARLSEN, B & EF GARDNER. 1994. *Stanford diagnostic reading test*. San Francisco: Harcourt Brace.
- KARLSEN, B & EF GARDNER. 1995. *Stanford Diagnostic Reading Test* (4th ed). New York: The Psychological Corporation.
- KARLSEN, B, R MADDEN & EF GARDNER. 1986. *Stanford Diagnostic Reading Test: Directions for Administering*. Brown Level. (3rd ed). New York: The Psychological Corporation.

- LIPKA, O & LS SIEGEL. 2012. The development of reading comprehension skills in children learning English as a second language. *Reading and Writing*, 25(8):1873 – 1898.
- LOW, PB & LS SIEGEL. 2005. A comparison of the cognitive processes underlying reading comprehension in native English and ESL speakers. *Written Language and Literacy*, 8(2): 207-231.
- LUKHELE, B. 2009. Exploring relationships between reading attitudes, reading ability and academic performance among teacher trainees in Swaziland. Unpublished Electronic Dissertation. Pretoria: University of South Africa.
- 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):54-67.
- MENKEN, K & T KLEYN. 2010. The long-term impact of subtractive schooling in the educational experiences of secondary English language learners. *International Journal of Bilingual Education and Bilingualism*, 13(4):399-417.
- MILLER, N. 1984. Language Problems and Bilingual Children. In N Miller (Ed), *Bilingualism and Language Disability: assessment and remediation*. London: Croom Helm. 81-103.
- MURPHY, KR & CO DAVIDSHOFER. 2004. *Psychological Testing: Principles and Applications* (6th ed). New Jersey: Pearson Prentice Hall.
- MURPHY, KR & CO DAVIDSHOFER. 1998. *Psychological Testing: Principles and Applications* (4th ed). New Jersey: Prentice Hall.
- NEALE, MD. 1989. *Neale Analysis of Reading Ability – Revised British Edition*. Manual. Oxford: NFER – Nelson Publishing Company.
- OWEN, K. 1998. *The Role of Psychological Tests in Education in South Africa: issues, controversies and benefits*. Pretoria, RSA: Human Sciences Research Council.
- POMPLUN, M & MH OMAR, 2001. Do reading passages about war provide factorially invariant scores for men and women? *Applied Measurement in Education*, 14(2):171-189.
- PRETORIUS, EJ. 2000. 'What they can't read will hurt them': reading and academic achievement. *Innovation*, 2:33-41. Editorial Committee, c/o University Library, University of Natal.
- PRETORIUS, EJ. 2010. Issues of complexity in reading: Putting Occam's razor aside for now. *South African Linguistics and Language Studies*, 28(4):339-356.
- PRETORIUS, EJ & S CUMMINS. 2010. Do the rich get richer and the poor poorer? The effects of an intervention programme on reading in the home and school language in a high poverty multilingual context. *International Journal of Educational Development*, 30:67-76.
- RICKETTS, J, K NATION & VM BISHOP. 2007. Vocabulary is important for some but not all reading skills. *Scientific Studies of Reading*, 11(3):235-257.
- RUTTER, M, A CASPI, DM FERGUSSON, LJ HORWOOD, R GOODMAN, B MAUGHAN, TE MOFFITT & J CAROLL. 2004. Gender differences in reading difficulties: Findings from four epidemiology studies. *Journal of the American Medical Association*, 291:2007-2012.

- SAN DIEGO UNIFIED SCHOOL DISTRICT: Standards, Assessment, and Accountability Division. Stanford Diagnostic Reading Test (SDRT) Summaries 2003-04. Available <http://studata.sandi.net/research/sdrt/index.asp>
- SEDL READING RESOURCES. The Stanford Diagnostic Reading Test (4th ed). Available <http://www.sedl.org/cgi-bin/mysql/rad.cgi?searchid=228>
- SHUTTLEWORTH-EDWARDS, AB, RD KEMP, AL RUST, JGL MUIRHEAD, NP HARTMAN & SE RADLOFF. 2004. Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS III test performance. *Journal of Clinical and Experimental Neuropsychology*, 26(7):903-920.
- SHUTTLEWORTH-JORDAN, AB. 1996. On not reinventing the wheel: A clinical perspective on culturally relevant test usage in South Africa. *South African Journal of Psychology*, 26 (2):96–102.
- SHUTTLEWORTH-JORDAN, AB. 1995. Against mutton masquerading as lamb: additional commentary on the interpretation and misinterpretation of the SA-WAIS. *South African Journal of Psychology*, 25:244–246.
- SKUY, M, E SCHUTTE, P FRIDJHON & S O’CARROLL. 2001. Suitability of Published Neuropsychological Test Norms for Urban African Secondary School Students in South Africa. *Personality and Individual Differences*, 3:1413-1425.
- WINNETT, S. L. 2008. The reading abilities of first and second language English learners in Grade 8, Grade 9 and Grade 10. Unpublished MA dissertation. University of the Witwatersrand.
- WOLFAARDT, JB & G ROODT. 2005. Basic Concepts. In Foxcroft, C & G Roodt (Eds), *An Introduction to Psychological Assessment in the South African context* (2nd ed). Cape Town: Oxford University Press. 25-45.

BIOGRAPHICAL NOTE

Dr Broom is a senior lecturer in the Psychology Department at the University of the Witwatersrand. She trained as an educator and worked as a teacher before returning to University to complete her PhD. She is interested in research into reading in multilingual educational contexts.
E-mail address: yvonne.broom@wits.ac.za

Deborah Diane Jewson is an educational psychologist in private practice. She is also a supervising psychologist at the Bellavista (Supporting Education through Evaluation and Knowledge) S.E.E.K assessment unit. Her fields of interest include assessment of reading and reading interventions in the South African context.