Investigating the Psychometric Properties of a South African Adaptation of the Boston Naming Test: Evidence for diagnostic validity from a Memory Clinic population

Lauren Baerecke

BRCLAU004

A minor dissertation submitted in partial fulfillment of the requirements for the award of the degree of Master of Arts (Psychological Research)

Department of Psychology

Faculty of the Humanities

University of Cape Town

2013

COMPULSORY DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature: Date: 13 September 2013

Acknowledgements

I would like to take this opportunity to express my appreciation to my supervisor, Kevin Thomas, for his support and guidance on this project. I would also like to thank Dr. Sebastiana Kalula and Sonya Hendricks for their help in accessing patient files at the Memory Clinic and accommodating my presence at the Institute for Aging in Africa. I wish to express my appreciation to the management and staff of the housing complex from which a number of participants for this study were drawn, for their interest in the project and their hospitality during the times I spent there. In addition, my sincerest thanks go to the men and women who gave of their time to participate in this research, and without whom this project would not have been possible. Lastly, I wish to thank my family and friends for their encouragement, support, and patience.

Scholarships: Thanks to the University of Cape Town University Research Scholarship for supporting this project.

Table of Contents

List of Figures

List of Tables

Abstract

The Boston Naming Test (BNT) is a popular confrontation naming test that is frequently used in the detection of naming deficits in Alzheimer's disease (AD). However, the test may not be appropriate when used outside of North America due to the influence of varying word frequency and familiarity between different cultures and languages. This study investigated the diagnostic validity of a South African 15-item adaption of the BNT (the BNT-SA-SF) in a Cape Town memory clinic population of patients with dementia and healthy, community-dwelling control participants. Between-groups comparisons, receiver operating characteristic (ROC) analyses, and other diagnostic efficiency statistics were used to assess the test's discriminative capacity between patients with AD *(n = 46)*, patients with other types of dementia *(n = 23)*, and controls *(n = 51)*, matched on key demographic variables. The AD group performed worse than patients with other types of dementia and controls on the BNT-SA-SF, and patients with other types of dementia scored more poorly than controls. The test showed the most significant discriminative capacity between patients with AD and controls, however. A general linear model examining the effects of sociodemographic variables on test performance found that BNT-SA-SF performance was not significantly affected by the sociodemographic characteristics of participants, including age, education, language, or socio-economic status, with the exception that men appear to achieve higher scores than women. Further, an item analysis identified a number of problematic items and suggestions are made concerning how to deal with these in future studies. Preliminary normative data stratified by sex and education are presented. Results support the clinical utility of the BNT-SA-SF as a screening test to aid in the diagnosis of AD from normal aging with older adults in South Africa. This study is a valuable step forward in the ongoing attempt to provide culturally appropriate and valid neuropsychological tests and norms for clinical and research purposes in South Africa. Future studies should examine the functioning of the test in larger samples, representative of the other major population and language groups in South Africa.

Introduction

The Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983, 2001) is a popular visual confrontation naming test that is widely used in the assessment of patients with a variety of neurological and cognitive impairments (Strauss, Sherman, & Spreen, 2006). Numerous studies (de la Plata et al., 2009; Pedraza, Graff-Radford et al., 2009; Worrall, Yiu, Hickson, & Barnett, 1995) have examined the issues that arise when using the test on culturally and linguistically diverse populations. The identification, in those and other studies, of cultural bias within the test has led to the modification of the BNT for use in languages and cultures different from those in which the test was developed.

The BNT and its various short forms are popular among South African clinicians and researchers. Despite this popularity, and the growing interest in cross-cultural neuropsychology in South Africa, there is little work involving the BNT in this country.

A Brief Introduction to the BNT

The BNT is the most widely used test of confrontation naming ability (Rabin, Barr, & Burton, 2005). Confrontation naming is defined as "the ability to pull out the correct word at will" (Lezak, Howieson, & Loring, 2004, p. 511). An early experimental version of the BNT consisted of 85 items (Goodglass, Kaplan, & Weintraub, 1976). Two revisions followed (Kaplan et al., 1983, 2001); each consisted of 60 items derived from the original pool of 85. Although the same 60 items comprise the first and second editions of the instrument, the latter (the BNT-2) differs from the former in that it includes a 15-item short form, a multiple-choice section, error codes for incorrect responses, and a discontinuation rule of eight consecutive failures instead of six. The BNT-2 is also included in the revised Boston Diagnostic Aphasia Examination (BDAE; Goodglass, Kaplan, & Barresi, 2001). Normative data are included in the BNT-2 record booklet for children aged 0-12 years and adults aged 18-79 years, split across a number of age bands.

In its current form, the BNT-2 consists of 60 black-and-white line drawings that are presented in ascending order of difficulty, from common items that are easier to name, such as *bed* and *toothbrush*, to less familiar items that are more difficult to name, such as *protractor* and *palette*. Standard administration begins with item 30 (see Strauss et al. (2006) for detailed administration and scoring guidelines). Credit is given for all previous items unless the examinee fails to produce a correct response for one of the first eight items. If the latter occurs, items are administered in reverse order until the examinee achieves eight consecutive correct responses. Standard administration dictates that an examinee is given 20

seconds to produce a correct spontaneous response. If the examinee's response shows misperception of the picture, a semantic cue is offered. Failing the production of a correct response following semantic cueing, a phonemic cue is offered. For instance, the semantic cue offered for the item *protractor* is 'it measures angles', and the phonemic cue for this item is 'it starts with the sound *pro*.' If the examinee does not produce a correct response after phonemic cueing, the examiner proceeds to the next item and returns to the failed item during administration of the multiple-choice section.

The multiple-choice section begins only after the examiner has completed the standard presentation described above. The examiner returns to each item the examinee has not named correctly, and the examinee is asked to select the correct choice (i.e., the word that describes the pictured object best) from four multiple-choice options read aloud by the examiner. All responses and the results of the multiple-choice questions are recorded in the test booklet, but full credit is given only to correct uncued responses and to correct answers in response to a semantic cue.

Development of BNT Short Forms

Researchers have developed a number of BNT short forms (Fastenau, Denburg, & Mauer, 1998; Huff, 1986; Lansing, Ivnik, Cullum, & Randolph, 1999; Mack, Freed, Williams, & Henderson, 1992; Morris et al., 1989; Nebreda et al., 2011; Saxton, Ratcliff et al., 2000; Tombaugh & Hubley, 1997; B. Williams, Mack, & Henderson, 1989). These shorter tests are useful for several reasons. Time constraints in clinical and research settings mean that administering the full 60-item test is impractical in some situations. Due to the number of items and cueing format of the test, administration can be lengthy and time-consuming. In other words, short forms aid in the rapid screening of patients. Reduced test time facilitates the assessment of patients with limited attention or motivation or those with severe neurological impairment who may become easily fatigued or frustrated (Jefferson et al., 2007). In addition, the development of equivalent short forms is useful in reducing practice effects in situations when multiple assessments of a patient are needed (Fastenau et al., 1998).

Reports on the psychometric properties of the numerous short forms vary. Studies of 30-item forms have reported good reliability and validity coefficients, and high correlations with the full test (Fisher, Tierney, Snow, & Szalai, 1999; Franzen, Haut, Rankin, & Keefover, 1995; Graves, Bezeau, Fogarty, & Blair, 2004; Saxton, Ratcliff, et al., 2000; Tombaugh & Hubley, 1997). Others have found that 15-item versions have good psychometric properties

(del Toro et al., 2011; Ferraro & Barth, 2003; Franzen et al., 1995), and that performance on them correlates well with performance on the full version (Calero, Arnedo, Navarro, Ruiz-Pedrosa, & Carnero, 2002; Mack et al., 1992; Tombaugh & Hubley, 1997). Generally, the psychometric properties of 30-item versions are slightly superior to those of the 15-item versions. After assessing eight short forms, Tombaugh and Hubley (1997) advised that 30-item versions are preferable to 15-item tests. Below, I review in more depth some of the most notable BNT short forms.

Mack et al. (1992) developed four equivalent 15-item versions of the BNT by dividing the 60 items of the original test into four 15-item groups, each reflecting the full range of content of the original test. These four tests, named Short Form (SF) 1-4, correlate strongly with each other, and each extrapolates well to the full BNT by multiplying the score out of 15 by 4 (Franzen et al., 1995; Mack et al., 1992). Tombaugh and Hubley (1997) report good correlations with the full test, with r ranging from .72 to .82, but lower internal consistency, with Cronbach's alpha ranging from .31 to .49. The highest correlation and reliability coefficients reported are for the SF-4 and the lowest for the SF-1. Of the four Mack short forms, Fastenau et al. (1998) reported higher reliability and validity coefficients for the SF-3 and SF-4, and they therefore recommended the use of these two over the others, particularly in situations when repeated testing is required. The Mack SF-4 is the short form included with the officially licensed BNT kit sold by the test publisher that markets the instrument.

Another popular short form is the 30-item test derived empirically by B. Williams et al. (1989). Those authors studied three 30-item versions. The first version comprised of all the even-numbered items in the full test, the second comprised of all the odd-numbered items, and the third was derived empirically by choosing the 30 items that best discriminated between Alzheimer's disease (AD) patients and healthy controls. They reported good criterion-related validity, excellent internal consistency (.93 to .96), and strong correlations with the full BNT (.94 to .99) for all three 30-item versions. The empirically derived test has particularly strong psychometric properties. Cross-validation studies (Franzen et al., 1995; Tombaugh & Hubley, 1997) report that it has excellent reliability and that performance on it correlates well with performance on the full BNT. It also has good internal consistency, with values similar to that of the full BNT (r = .74 for the empirical test and .78 for the full test; Tombaugh & Hubley, 1997). Further, the empirically-derived test appears to have diagnostic accuracy in classifying impaired and healthy individuals in agreement with the full BNT (Franzen et al., 1995; Lansing et al., 1999).

The 15-item short form devised for the Consortium to Establish a Registry for Alzheimer's Disease (CERAD; Morris et al., 1989) is also widely used. However, the literature generally identifies this version as the least desirable short form (Franzen et al., 1995; Larrain & Cimino, 1998; Mack et al., 1992). Scores on the CERAD 15-item BNT have been shown to extrapolate poorly to scores on the full BNT (Hobson et al., 2011; Mack et al., 1992), and performance on it correlates poorly with performance on the full BNT ($r = .62$; Tombaugh & Hubley, 1997). It also has poor criterion-related validity compared to the full BNT (Larrain & Cimino, 1998), poor internal consistency compared to other 15-item short forms (Franzen et al., 1995; Tombaugh & Hubley, 1997), and a higher average item difficulty than the full BNT (Franzen et al., 1995).

Despite the apparent value of these abbreviated versions, the normative data available for them are limited, particularly when they are used in less educated and more culturally diverse samples. Without investigation of the psychometric properties and the provision of normative data, the clinical and research utility of these short forms is limited (Jefferson et al., 2007).

Clinical Findings

Confrontation naming is sensitive to indicators of various neurological and developmental disorders (Lezak et al., 2004). Thus, clinicians frequently assess this cognitive function as a component of a neuropsychological evaluation. The diagnostic information provided by the BNT, whether in full or short form, is frequently used in the assessment of naming deficits in patients of all ages and with different clinical pathologies.

The BNT is commonly used to assess language performance in patients with aphasia (Kohn & Goodglass, 1985; S. Williams & Canter, 1982) such as that frequently seen following a stroke (del Toro et al., 2011; Kendall, Rodriguez, Rosenbek, Conway, & Rothi, 2006). Its use in clinical studies includes, but is not limited to, assessing naming difficulties in patients with degenerative disorders such as multiple sclerosis (Lezak, Whitman, & Bourdette, 1990), in children with dyslexia (Scarborough, 1990), in surgical epilepsy patients (Busch, Frazier, Haggerty, & Kubu, 2005; Davies et al., 1998; Loring et al., 2008), and in people of all ages following mild head trauma (Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005; Yeates et al., 2002).

Confrontation naming in dementia. Naming impairment is a common feature of the language disorder seen in dementias of various etiologies. Many studies have identified word finding difficulties in dementia patients, without classifying according to etiology, using

various confrontation naming tests (Brouillette et al., 2011; Kirshner, Webb, & Kelly, 1984; Miller, Finney, Meador, & Loring, 2010; Rochford, 1971). For instance, Brouilette et al. (2011) demonstrated the ability of the Memory for Names Test, a 75-item confrontation naming test utilizing pictures of famous political and historical figures and celebrities, to discriminate normal aging from mild cognitive impairment (MCI) and MCI from dementia.

The ability to differentiate dementia from normal aging is frequently used as an indicator of the criterion-related validity of BNT short forms (Calero et al., 2002; de la Plata et al., 2009). For example, in Spain, Nebreda et al. (2011) found that a Spanish translation of the BNT successfully differentiated patients with dementia from patients with other diagnoses (e.g., depression and MCI) and controls.

The BNT in Alzheimer's disease (AD). Although naming impairment is a common language deficit reported in many types of dementia, it is particularly substantial in AD. Therefore, the BNT is frequently used in the neuropsychological detection of possible or probable AD. Despite some evidence that classification rates differ across various short forms, numerous studies have shown the effectiveness of both short and full versions of the BNT in distinguishing cognitively intact older adults from patients with AD. Patients with AD tend to make significantly more errors on the BNT than cognitively intact older adults (Balthazar, Cendes, & Damasceno, 2008; De Jager, Hogervorst, Combrink, & Budge, 2003; Knesevich, LaBarge, & Edwards, 1986; Lukatela, Malloy, Jenkins, & Cohen, 1998; B. Williams et al., 1989), with performance declining steadily as disease severity increases (Faber-Langendoen et al., 1988; Price et al., 1993). LaBarge, Balota, Storandt, and Smith (1992) found that patients with AD who were classified as 'mildly demented' scored significantly lower on the 60-item BNT than those who were classified as 'very mildly demented' and that both patient groups performed more poorly than healthy controls. Similarly, but using a 30-item test, Chosak Reiter (2000) found that severely demented AD patients scored significantly more poorly than moderately demented AD patients, who scored significantly more poorly than mildly demented AD patients.

Short forms have been found to demonstrate good criterion-related validity in that they are able to correctly discriminate AD patients from normal controls at rates comparable to the long form's classification (Calero et al., 2002; Fisher et al., 1999; Graves et al., 2004; Jefferson et al., 2007; Mack et al., 1992). For instance, using the full BNT, Lansing et al. (1999) showed a significant difference in total score between 325 AD patients and 719 elderly normal control participants. The test correctly classified 67.5% of the former and 84.4% of the latter; nine 15- and 30-item versions classified participants correctly at a similar

level. In fact, the discriminative capacity of their empirically created 15-item short form was slightly higher than that of the 60-item test.

A review of previously created short forms suggests that the B. Williams et al. (1989) 30-item test is the most successful in discriminating between AD and normal controls, having been well-validated for this purpose in a number of studies (Franzen et al., 1995; Graves et al., 2004; Lansing et al., 1999). However, a number of other short forms also demonstrate clinical utility. For instance, Mack et al. (1992) found that the four 15-item versions they developed, as well as three 30-item versions, successfully differentiated a sample of 26 AD patients from a sample of 26 healthy controls.

Whereas the Mack short forms appear to discriminate between AD patients and controls nearly as well as the 60-item test does (Lansing et al., 1999), there are conflicting reports of variable classification rates and high misclassification rates across the four tests and across other short forms (Franzen et al., 1995). Graves et al. (2004) found that 12 short forms, including two new short forms developed by the authors using the principles of item response theory (IRT), showed a variable rate of agreement with the full BNT in classifying patients with AD or a combination of AD and vascular dementia (VaD) as impaired. Some tests showed a high rate of agreement, such as the authors' 15-item version, the Mack SF-3, and the Mack SF-4. Other tests, however, performed poorly, such as the Mack SF-1, the Mack SF-2, the Lansing et al. (1999) 15-item, and one of the 30-item tests created by Saxton, Ratcliff et al. (2000).

The BNT in early AD. As increasing importance is placed on the early detection and prevention of AD, it is valuable if a neuropsychological test such as the BNT is able to distinguish the cognitive impairment in AD from other dementia types and normal aging, in the initial stages of the disease.

There is evidence that the BNT is able to detect naming deficits present in the early stages of AD (Chen et al., 2001; B. Williams et al., 1989). Prospective studies reveal that changes in confrontation naming are sometimes evident even in the preclinical stages of the disease. Jacobs et al. (1995) found that baseline scores on a 15-item BNT in healthy nondemented adults were significantly associated with subsequent diagnosis of AD. The authors concluded that assessment of confrontation naming might be useful in detecting AD in its earliest stage, even before functional impairment manifests. Similarly, Welsh, Butters, Hughes, Mohs, and Heyman (1992) found that the BNT short form was the only nonmemory subtest of the CERAD battery to show utility in distinguishing patients with early AD from

normal control participants. Others have found that the test discriminates very mildly demented AD patients from controls (Chosak Reiter, 2000; LaBarge et al., 1992)

Some evidence suggests, however, that the BNT is only able to detect naming deficits in moderate or severe AD (Bayles & Tomoeda, 1983). For instance, in a Portuguese-speaking Brazilian sample, Bertolucci et al. (2001) did not find a significant difference between controls and individuals with mild AD (defined as a Clinical Dementia Rating (CDR) score of 1) but did find a significant difference between these two groups and individuals with more severe AD (defined as a CDR score of 2) on the 15-item CERAD BNT. Testa et al. (2004) found that impairment on the BNT is common in moderate or severe AD but poor performance on measures of delayed recall and category fluency is more prevalent in early AD. The authors conclude that the impairment on the BNT is "neither necessary nor sufficient" for establishing the diagnosis of AD early in the disease process (Testa et al., 2004, p. 511).

Thus, although it is clear that naming impairment is a feature of AD, the evidence as to whether this deficit becomes apparent in the early stages of the disease is mixed. The BNT may be sensitive to cognitive changes in early AD, but any deficits in test performance will most likely be subtle in the prodromal stages of the dementia and ubiquitous only in the later stages.

Using the BNT to distinguish AD from other types of dementia. Although empirical comparisons of naming deficits in different types of dementia are quite rare in the published literature, there is some evidence that assessment of confrontation naming using the BNT can facilitate the differential diagnosis of AD from other types of dementia.

Differences in BNT performance between AD and vascular dementia. In terms of BNT-based differential diagnoses among dementias, the comparison between vascular dementia (VaD) and AD has been the most widely researched. Although patients with VaD also exhibit naming difficulties compared to controls (Chosak Reiter, 2000), AD patients tend to make significantly more naming errors than patients with VaD (Barr, Benedict, Tune, & Brandt, 1992; Lukatela et al., 1998; Schmidtke & Hüll, 2002; Villardita, 1993). De Jager et al. (2003) identified that the BNT was one of four tests that were able to discriminate between patients with AD and patients with cerebrovascular dementia (CVD) in a neuropsychological battery designed to help distinguish between dementia, mild cognitive impairment (MCI), and controls. In this study, patients with AD performed more poorly than patients with CVD, which included those diagnosed with vascular cognitive impairment or VaD, and both groups of patients performed more poorly than controls.

There are some conflicting findings in this area of the literature, however. For instance, Baillon et al. (2003) reported similar performance on the BNT across the two diagnoses. Mathias and Burke (2009) pointed to the difficulty in differentiating between AD and VaD using the BNT, particularly as recent research has disputed the traditional understanding that these two dementias result from different etiologies. Using a meta-analysis, these authors found that picture naming tests, including the BNT, did not discriminate successfully between AD and VaD. They report a weighted mean effect size of -0.4 for tests of picture naming.

Differences in BNT performance between AD and mixed Alzheimer's and vascular dementia. Studies comparing the neuropsychological test profiles of individuals with AD and individuals with mixed AD/VaD generally report that individuals with these two types of dementia perform similarly on the BNT. Miller et al. (2010) found no difference between patients with mixed AD/VaD and those with probable AD on a 15-item BNT and on another visual confrontation naming task. There were no between-group differences in overall accuracy, or the number of items correct after semantic or phonemic cueing. This pattern of data is consistent with that reported by Schmidtke and Hüll (2002), who also found no difference in BNT performance between patients with AD and those with mixed AD/subcortical VaD on the BNT short form in the German version of the CERAD battery.

Differences in BNT performance between AD and frontotemporal dementia. A number of studies report similar performance on the BNT by FTD and AD patients (Grossman et al., 2004; Mendez et al., 1996). Others report that FTD patients perform similarly to controls (Pachana, Boone, Miller, Cummings, & Berman, 1996), suggesting that confrontation naming is not impaired in FTD. It appears, however, that certain FTD subtypes are particularly impaired with regards to confrontation naming. Specifically, patients with semantic dementia (SD), the temporal variant of FTD, appear to be the most impaired on naming tasks compared to patients with other FTD subtypes and to patients with AD (Grossman et al., 2004; Kramer et al., 2003). Diehl et al. (2005) reported the results of a logistic regression model which showed that the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) and a 15-item BNT correctly classified 96.3% of patients with either SD or AD. Interestingly, a combination of a low MMSE score and a high BNT score predicted AD in this sample.

Differences in BNT performance between AD and other types of dementia. There is mixed evidence regarding the differences in confrontation naming ability between individuals with AD and those with types of dementia other than those mentioned above. AD patients

have been shown to perform more poorly than people with alcohol-related dementia on the BNT, who in turn perform similarly to controls (Saxton, Munro, Butters, Schramke, & McNeil, 2000), which suggests that BNT performance is not impaired in alcohol-related dementia. Although some studies report lower scores on the BNT in Lewy Body dementia than in AD (Ferman et al., 2006; V. Williams et al., 2007), Noe et al. (2004) found no difference in performance between patients with AD, Lewy Body dementia, or Parkinson's disease with dementia on a 15-item BNT short form. Whereas Hodges, Salmon, and Butters (1991) found that AD patients were significantly more impaired than Huntington's disease patients on the BNT, Bayles and Tomoeda (1983) found that a confrontation naming task did not discriminate between patients with mild AD, mild Parkinson's disease with dementia, mild Huntington's disease, or multi-infarct dementia, and healthy controls. Bayles and Tomoeda (1983) did report, however, that patients with moderate AD made significantly more errors than patients with moderate Parkinson's disease with dementia, patients with moderate Huntington's disease, and healthy controls. This finding suggests that confrontation naming is not impaired in mild dementia cases, even AD; although, it must be noted that the authors in this study did not account for patient level of education, age, or sex in interpreting BNT performance.

Cognitive Processes Underlying Confrontation Naming

The BNT and other picture naming tasks are deceptively simple; confrontation naming is, in fact, a complex neurocognitive process. Successful performance requires the integration of several underlying cognitive processes and the participation of numerous neural substrates. There is a large body of literature on these mechanisms, and a number of models of language processing have been generated to account for the process by which individuals name objects (see e.g., Glaser, 1992; Paivio, 1991; Seymour, 1973; see Levelt, 1999 for a review of models of word production). Generally, the literature identifies at least three main processes involved in picture naming (DeLeon et al., 2007; Johnson, Paivio, & Clark, 1996; Paivio, Clark, Digdon, & Bons, 1989):

1. The first step is perceptual analysis of the stimulus and involves visually identifying and recognizing the object.
2. The second step is name activation and involves activation of the lexical-semantic representation of the object in memory.

3. The third step is response generation and involves phonological activation and subsequent motor execution of the word (i.e., planning and implementing the articulation of the word).

The nature of the naming deficit in Alzheimer's disease. To examine where the naming deficit in AD lies in the above-mentioned processes, numerous studies have analyzed the types of errors made on the BNT and other confrontation naming tasks by people with AD. These studies have analyzed error quality, grouping them into categories such as visuo-perceptual errors, semantically-related errors, and omissions or unrelated responses. Corresponding to the steps involved in picture naming outlined above, naming errors in patients with AD are generally attributed to step one, perceptual impairment, or step two, lexical-semantic impairment (Barresi, Nicholas, Tabor Connor, Obler, & Albert, 2000; Nebes, 1989). The naming deficit in AD is generally not attributed to the third broad step, phonological activation and subsequent motor execution of the word (Hodges et al., 1991).

Perceptual impairment. Early studies in the field suggested the naming deficit in dementia could be attributed to impairment in visual identification or recognition. These studies showed, mainly by manipulating the visual quality of the object, that the naming errors made by dementia patients were predominantly perceptual in nature (Kirshner et al., 1984; Shuttleworth & Huber, 1988). Using an 8-item naming test, Rochford (1971) showed that 55% of the errors made by a sample of patients with dementia indicated misperception, a far greater rate than in a group of dysphasic patients with focal lesions. When asked to name familiar, visible parts of their own bodies (ear, elbow, etc.), only 2 of 17 dementia patients made any errors, compared to 16 of 17 who made errors on the object-naming test. The author took this improved naming performance with these visually familiar objects as evidence that the problem lay with recognition.

The finding that AD patients tend to make more semantic errors, rather than other errors, on object naming tasks (Bayles & Tomoeda, 1983; Martin & Fedio, 1983) led to suggestions that impairment of semantic knowledge, rather than visuo-perceptual impairment, explained the naming deficit seen in AD. Thus, the hypothesis that naming difficulties in AD are primarily due to impaired visual perception has generally been disregarded in recent literature. However, there is still debate as to the predominant disturbance behind the naming deficit.

Lexical-semantic impairment. Many studies support impairment within the lexical-semantic sphere as the primary cause of the naming deficit in AD (Bayles & Tomoeda, 1983; Chertkow, Bub, & Seidenberg, 1989; Garrard, Lambon Ralph, Patterson, Pratt, & Hodges,

2005; Huff, Corkin, & Growdon, 1986; S. Rogers & Friedman, 2008; T. Rogers, Ivanoiu, Patterson, & Hodges, 2006). However, it is not clear whether the nature of this disruption is that (a) the content and organization of the semantic information itself is disturbed or lost altogether (the semantic degradation view), or (b) the semantic structure remains intact but access to it is compromised (the lexical access view; Barresi et al., 2000; Gainotti, 1987; Nebes, 1989).

Semantic degradation hypothesis. According to this hypothesis, the impaired naming performance in AD reflects a fundamental disruption or loss of stored semantic information (Nebes, 1989). Studies supporting the semantic degradation view point to consistent errors across two or more administrations of the test (Gainotti, 1987) or across various tests of semantic memory using the same items (Chertkow & Bub, 1990; Hodges, Salmon, & Butters, 1992). For example, Henderson, Mack, Freed, Kempler, and Andersen (1990) demonstrated an error consistency of 80% on the BNT on two administrations, 6 months apart, in a sample of patients with AD. This pattern of data suggested to the authors that naming impairment in AD is, at least in part, due to a loss of semantic information. After finding consistency between errors on the same items on a confrontation naming test and a name recognition test, Huff et al. (1986) drew a similar conclusion that the semantic information about these specific items and their names were lost in AD. In addition, the finding of a frequency effect, in which AD patients display a disproportionately high error rate for naming low frequency or less familiar items compared to controls, is also seen as confirmation of the semantic degradation hypothesis (Hodges et al., 1992).

Lexical access hypothesis. According to this hypothesis, the impaired naming performance in AD reflects a disruption in accessing the lexical-semantic field, i.e., in accessing the specific name of the object. In other words, the semantic representation of the object is preserved but there is a problem in gaining access to, or in retrieving from, the internal lexicon (Nebes, 1989). Studies supporting the lexical access view generally point to naming errors being inconsistent over time (Gainotti, 1987). Instances where individuals fail to name an item spontaneously but benefit from a phonemic cue have also been used to support the lexical access hypothesis (Martin & Fedio, 1983; Neils, Brennan, Cole, Boller, & Gerdeman, 1988). For instance, Balthazar et al. (2008) found that patients with mild AD made significantly more errors than patients with mild cognitive impairment (MCI) and than healthy controls after semantic cueing but that there were no significant between-group differences after consideration of scores following the administration of phonemic cues. Such

data are considered evidence that the individual's semantic knowledge about the item is intact but that they struggle to access the word.

Multiple underlying deficits. It is unlikely that a single cognitive deficit can account for the naming impairment in AD (Henderson et al., 1990). Some authors have argued that multiple underlying cognitive mechanisms are impaired within each individual. Recent evidence that AD patients' semantic memory is degraded, but only partially, and their impairment on semantic memory tasks such as picture naming is primarily due to deficient explicit retrieval, suggests a dual impairment theory (S. Rogers & Friedman, 2008). Other authors have argued that the pattern of error types changes as a function of dementia severity (Bayles & Tomoeda, 1983; Kim, Kim, & Na, 1997). Studies have found that errors reflecting a loss of lexical information, such as producing a word semantically related to or of the same category of the target word, occur early in the disease process (Barbarotto, Capitani, Jori, Laiacona, & Molinari, 1998; LaBarge et al., 1992). In contrast, the same studies have found that errors reflecting a loss of semantic information, such as empty responses (e.g., 'I don't know') or responses that are unrelated to the target word in any way, occur more frequently as the disease progresses. These error patterns indicate that the naming deficit in AD is progressively semantic in nature; lexical access is impaired in early AD, but the semantic system breaks down as AD progresses from mild to more advanced stages of the disease.

One explanation for the mixed findings in this area is that different studies examining BNT performance in dementia use different error classification systems. For instance, Shuttleworth and Huber (1988) interpreted what they called 'failure to verbalize' responses, most of which were 'I don't know' responses, as most likely being perceptual in nature. In contrast, LaBarge et al. (1992) classified 'no-content errors', including empty phrases such as 'I don't know' or 'can't think of it', as reflecting a loss of semantic information. The authors in the latter study justify this by saying such responses indicate there is insufficient semantic information to produce a response that is related to the target word. Errors on the BNT are often ambiguous in nature, which perhaps indicates that they could be produced by failure of multiple cognitive processes, and that it would therefore be mistaken to attribute them to a single cause (Chosak Reiter, 2000; Lukatela et al., 1998).

In summary, although simplified and not absolute, qualitative analysis of naming errors has proved an important tool for investigating the functional deficit underlying the naming difficulty in AD. It is not clear whether the well-established naming deficit in AD can be attributed to a breakdown in accessing intact semantic information, to a disruption or loss of the semantic information itself, or to a combination of the two mechanisms (Mitrushina,

Boone, Razani, & D'Elia, 2005). Nonetheless, the BNT has been widely used as a diagnostic tool in distinguishing individuals with AD from nondemented individuals and from those with other types of dementia, as well in assessing language function in a variety of other disorders. Due to the widespread use of the test in detecting even mild naming impairments, many authors have remarked on the importance of investigating non-neurological factors that may affect BNT performance.

Moderators of BNT Performance

Confirming the clinical utility of the BNT is, to some degree, dependent on identifying non-neurological factors that may affect test performance. Although sensitive to acquired naming deficits, BNT performance is also influenced by demographic factors. The literature identifies at least six such factors: age, education, vocabulary, sex, language, and culture (although the latter two are closely linked).

Age and BNT performance. There is evidence that age affects confrontation naming. Numerous empirical studies on the effects of 'normal' aging on BNT performance, using the full version or short forms of the test, report that naming ability decreases with advancing chronological age (Cruice, Worrall, & Hickson, 2000; Kimbarow, Vangel, & Lichtenberg, 1996; Lucas et al., 2005; Neils et al., 1995; Saxton, Ratcliff et al., 2000; Worrall et al., 1995; Zec, Burkett, Markwell, & Larsen, 2007a, 2007b). These performance deficits are particular evident past the seventh decade of life (Au et al., 1995; Mitrushina & Satz, 1989; M. Nicholas, Obler, Albert, & Goodglass, 1985; L. Welch, Doineau, Johnson, & King, 1996; Zec, Markwell, Burkett, & Larsen, 2005), although others have reported the significant decline is only observed after the age of 80 years (Kent & Luszcz, 2002).

One of the largest studies of its kind ($N = 1017$ healthy individuals, aged between 50 and 99 years), reported significantly poorer mean scores, as well as greater variability in scores, in successively older age groups (Zec et al., 2007a). The authors emphasized, however, that although these declines were statistically significant, they represented only a modest decline in ability, and that confrontation naming remains *relatively* intact in normal aging.

This position is supported by findings on the magnitude of the correlation between BNT scores and age. Van Gorp, Satz, Kiersch, and Henry (1986) report an *r* of -.33 that is consistent with other studies (Kimbarow et al., 1996; MacKay, Connor, & Storandt, 2005; Neils et al., 1995; L. Welch et al., 1996). In addition, the increased variability with age reported by Zec et al. (2007a) is congruent with findings from other studies (LaBarge,

Edwards, & Knesevich, 1986; L. Welch et al., 1996), which suggests that naming decline is not a universal nor an inevitable phenomenon of aging. Otherwise stated, only some, not all, older adults have naming difficulties (Goulet & Ska, 1994).

Of interest here is that a narrative review of 25 picture-naming studies found that an age-related decline in picture naming is an inconsistent finding (Goulet & Ska, 1994). A number of studies have not found significant effects of age on BNT performance (L. Nicholas, Brookshire, Maclennan, Schumacher, & Porrazzo, 1989; Tombaugh & Hubley, 1997) and others have found that older adults actually perform better on the BNT than younger adults (Farmer, 1990; Schmitter-Edgecombe, Vesneski, & Jones, 2000).

These inconsistent findings may be the result of variability in research design and sample characteristics. For example, using longitudinal analysis, Cruice et al. (2000) found that naming ability did not decline with chronological age in elderly adults over a 4-year period. However, cross-sectional analysis of the same data revealed a significant, albeit weak (r_s= -0.21), correlation between age and BNT performance. Similar varying findings are reported elsewhere. For instance, Connor, Spiro, Obler, and Albert (2004) reported a larger age-related decline in BNT performance in their cross-sectional study than in their longitudinal study. Other longitudinal studies have found that naming ability does (Au et al., 1995) and does not (Mitrushina & Satz, 1995) show age-related decline.

A possible reason for these conflicting results may be due to cohort effects in cross-sectional studies and possible attrition bias or practice effects in the longitudinal studies (Cruice et al., 2000; Zec et al., 2007a). Taking this into consideration, Zec et al. (2007a) estimated the actual decline from the fifth to eight decade of life to be 2-3 words on the 60-item BNT. Thus, for the most part, the decline in naming ability with increasing age appears to be modest at best.

Education and BNT performance. An individual's confrontation naming ability is not only affected by age; level of education (i.e., number of years of formal education completed successfully) has been identified as one of the primary demographic factors that moderate BNT performance. Many studies report lower mean scores associated with lower levels of education in various BNT long and short forms (Jefferson et al., 2007; Kent & Luszcz, 2002; Kimbarow et al., 1996; Lansing et al., 1999; Neils et al., 1995; L. Nicholas et al., 1989; Saxton, Ratcliff et al., 2000; Tombaugh & Hubley, 1997; Weintraub et al., 2009; L. Welch et al., 1996; Worrall et al., 1995; Zec et al., 2007a). Findings on the magnitude of the correlation between education and BNT score vary from strong (.52; Hawkins et al., 1993), to

moderate (.40; Kimbarow et al., 1996), to weaker (.31 (Zec et al., 2007a) and .20 (Cruice et al., 2000)).

The relationship between BNT performance and education does appear to be stronger at lower levels of education, however. A number of studies have reported larger differences between participants who have completed high school and those who have not than between those who have completed high school and those who have some form of tertiary education (Hawkins & Bender, 2002; Le Dorze & Durocher, 1992; Neils et al., 1995). For example, Zec et al. (2007a) report that the greatest difference in BNT score exists between those with < 12 years of education and those with 12 or more years of education, but that those with 12 years and > 12 years of education perform similarly.

Furthermore, there may be an interaction between age and education. Education may mediate the relationship between age and BNT performance such that higher education levels may postpone the deleterious effects of aging to after the age of 80 years, compared to 70 years for those with < 12 years of education (L. Welch et al., 1996), or may even eliminate them altogether (Farmer, 1990). Neils et al. (1995) found that BNT scores for those with > 12 years of education were not affected by age in a sample of 323 elderly adults over the age of 65.

Although the evidence reviewed above shows strong support for the relationship between BNT performance and education, there are some studies that have not found a significant relationship (Au et al., 1995; Farmer, 1990; Fastenau et al., 1998; Hall, Vo, Johnson, Wiechmann, & O'Bryant, 2012; Heaton, Avitable, Grant, & Matthews, 1999; LaBarge et al., 1986; Mitrushina & Satz, 1995). The lack of significant findings in these studies may be due to small sample sizes (e.g., LaBarge et al. (1986) had a sample size of 58) and limited ranges of education (e.g., in Hall et al. (2012) the mean education level across all groups was > 14 years) in some of the studies. Most BNT normative studies, particularly earlier ones, have disproportionate representation of individuals with high levels of education, thus failing to adequately represent those with lower levels of education (Hawkins & Bender, 2002). As BNT scores are insensitive to differences at high levels of naming ability and individuals with a high level of education perform well on the BNT, the reported lack of a significant association between education and BNT performance in well-educated samples is perhaps unsurprising (Fastenau, 1998; Hawkins & Bender, 2002).

Vocabulary and BNT performance. Research also indicates those with better vocabulary achieve higher BNT scores. After reviewing the relationship between education, vocabulary, and BNT performance, Hawkins and Bender (2002) proposed that education acts

as a proxy for vocabulary. The authors stated that in normal healthy adults, premorbid vocabulary or verbal intelligence is often more strongly related to BNT score than level of education. Vocabulary would thus provide a better basis for norm stratification were it not for the fact that performance on vocabulary tests was also impaired by the pathology that results in poor BNT performance (i.e., dysnomia), whereas education is not affected by such factors.

A number of studies have pointed to verbal intelligence, specifically as measured by the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) Vocabulary Subtest, as being related to BNT performance (Thompson & Heaton, 1989; Tombaugh & Hubley, 1997; Wilkins, Hamby, & Thompson, 1996). Killgore and Adams (1999) found a correlation of .65 between WAIS-R Vocabulary Subtest score and BNT performance in a neurologically intact clinical sample; this measure of verbal intelligence accounted for 42% of the variance in this study. Tombaugh and Hubley (1997) reported a similar correlation (r = .53) between WAIS-R Vocabulary Subtest score and BNT performance. Wilkins et al. (1996) recommended that the norms included in the BNT test manual not be used for individuals with Vocabulary scaled scores less than or equal to 7.

In the absence of a verbal IQ score, reading vocabulary tests may be useful (Strauss et al., 2006). Hawkins et al. (1993) reported a correlation of .83 between BNT score and reading vocabulary, as measured by the Gates-MacGinite Reading Vocabulary Test (G-MRVT; MacGinite & MacGinite, 1989), in schizophrenic and bipolar patients and normal controls. In that sample, the correlation between BNT score and education was .52.

In summary, individuals with limited education or poor vocabulary skills may perform below widely accepted cut-off scores for impairment on the BNT. Hawkins and Bender (2002, p. 1143) conclude that "BNT norms should be finely stratified by education. Whenever possible, the clinical interpretation of BNT scores should be further moderated by estimations of premorbid vocabulary." Failure to recognize the influence of these variables may result in false positive diagnoses.

Sex and BNT performance. Research findings on sex differences in BNT performance have been inconsistent. Most studies have not found significant differences in the performance of men and women, whether using the 60-item administration or short forms (Cruice et al., 2000; Fastenau et al., 1998; Fastenau, 1998; Kent & Luszcz, 2002; LaBarge et al., 1986; Ross, Lichtenberg, & Christensen, 1995; Saxton, Ratcliff et al., 2000; Zec et al., 2007a). Some studies have, however, reported differences in favor of men in older samples (Lansing et al., 1999; Randolph, Lansing, Ivnik, Cullum, & Hermann, 1999; Tombaugh & Hubley, 1997; L. Welch et al., 1996). Recently, Hall et al. (2012) reported that men

performed significantly better than women, even after controlling for age, education, cardiovascular health, and IQ, in both cognitively intact (*n* = 468 normal controls) and cognitively impaired (*n* = 153 patients with probable AD) elderly samples. They concluded that sex is an important characteristic to account for when evaluating BNT performance.

Studies that have found significant sex differences in BNT performance have offered a number of explanations to account for this effect. One explanation is that significant sex difference in BNT performance may be attributable to an interaction between age and education. Cohort or generational effects may exist in older samples, within which men generally received a higher level of education than women (Zec et al., 2007b). L. Welch et al. (1996) noted that the older adults in their study completed their formal education more than 40 years prior to being tested, and at this time men often received a higher level of education than women. Jefferson et al. (2007) also found that sex differences are secondary to education differences between men and women. Another explanation is that the BNT contains a disproportionate number of male-biased items, which accounts for the significant effect of sex. Men have been found to perform significantly better than women on items such as *dart*, *canoe*, *muzzle*, *compass*, *sphinx*, and *protractor* (Randolph et al., 1999).

Although the effect of sex is equivocal, the effects of age and education on BNT performance appear more robust. Thus, sociodemographic factors may affect BNT performance, and clinicians and researchers alike need to be cautious in attributing poor performance to neurological impairment without considering these variables. Similar findings regarding the effects of age, education, and sex have been replicated across numerous countries and within different cultures and languages, including Australia (Worrall et al., 1995), Brazil (Miotto, Sato, Lucia, Camargo, & Scaff, 2010), Greece (Simos, Kasselimis, & Mouzaki, 2011), Korea (Kim & Na, 1999), Spain (Peña-Casanova et al., 2009), and Sweden (Tallberg, 2005), in French-speaking Canadians in Quebec (Roberts & Doucet, 2011), and in Dutch-speaking Belgians (Mariën, Mampaey, Vervaet, Saerens, & De Deyn, 1998). However, caution is warranted when using the BNT in such culturally and linguistically diverse settings. Clinicians and researchers widely accept that such tests, and their associated normative datasets, cannot simply be appropriated from one culture or language and applied directly to another, without modification.

Language, culture, and BNT performance. BNT performance is also influenced by linguistic and cultural background. The BNT, along with many other popular neuropsychological tests, was developed and standardized for the assessment of English-speaking, North American individuals, and the test reflects this sociocultural context. The

BNT may not be appropriate when administered to individuals from populations that differ from that on which it was standardized due to varying word-frequency and familiarity between different languages and cultures and the resultant lack of cultural relevance of certain items to those test-takers (Cruice et al., 2000; de la Plata et al., 2009; Kim & Na, 1999; Worrall et al., 1995). The use of the BNT and accompanying normative data may lead clinicians and researchers to overestimate the level of cognitive impairment present in these individuals.

BNT performance by individuals from outside of North America (i.e., from populations with different cultural and linguistic backgrounds from those in which the test was originally developed and normed) differs significantly from the performance of English-speaking, North American individuals. Belgians (Storms, Saerens, & De Deyn, 2004), African-Caribbean residents of the United Kingdom (Stewart, Richards, Brayne, & Mann, 2001), and South Africans (Mendoça, 2010), among others, have been shown to perform more poorly than North American normative standards. This cross-cultural effect on BNT performance has not only been found in populations that do not have English as a primary language. Although English-speaking Canadians perform as well as Americans (Tombaugh & Hubley, 1997), in other English-speaking countries, such as New Zealand and Australia, BNT performance of normal healthy individuals is significantly worse than that of comparable North American individuals (Barker-Collo, 2001; Worrall et al., 1995).

Comparisons to normative data are one of the main sources of information on which clinicians base decisions about a patient's level of impairment. When tests are used cross-culturally, as in the manner described above, such normative comparisons may be misleading; many clinicians and researchers have pointed to the potential for misdiagnosis of naming deficits when the BNT is used in such a way (Barker-Collo, 2007). Poor performance can mistakenly be considered a deficit when in fact it reflects that the item is unknown due to language or cultural differences. Anyone administering the BNT to individuals outside of the US, even if those individuals have English as a home language, must proceed with caution, especially because the potential for culturally biased results is high if individuals are not familiar with the sociocultural context in which the test was developed.

The same note of caution applies when administering the test to different ethnic groups or minorities within the US. For instance, several studies report that white Americans perform significantly better than African-Americans on the BNT (Fillenbaum, Huber, & Taussig, 1997; Kimbarow et al., 1996; Pedraza, Graff-Radford et al., 2009; Ross et al., 1995; Whitfield et al., 2000). Similar findings have emerged from studies on Spanish-speaking

samples in the US (Taussig, Henderson, & Mack, 1992). It appears that, with regard to different ethnic groups in the US, those who are most acculturated to the dominant or mainstream culture score higher than those less familiar with that culture. For example, Touradji, Manly, Jacobs, and Stern (2001) found that nondemented white non-Hispanic elderly adults born in the US scored significantly higher than those born outside the US on a 15-item BNT short form, despite the foreign-born participants speaking English 'very well' and living in the US for an average of 57.3 years. Within the group of foreign-born individuals, the effect of years spent in the US approached significance in predicting BNT score, after age and education were accounted for.

Such between-group differences arise primarily because of item bias. Evidence for this point is provided by examination of differences in the patterns of errors made by examinees with different cultural background. The literature identifies certain items as being especially culturally loaded. For instance, in both younger and older non-North American samples, the items *beaver* and *pretzel* frequently produce naming errors at a significantly higher rate than the adjacent items that should have a similar level of difficulty (Barker-Collo, 2001, 2007; Mariën et al., 1998; Piguet et al., 2001; Worrall et al., 1995). In the creation of the Korean version of the BNT, Kim and Na (1999) noted that *abacus*, one of the most difficult items on the test, would be easy for Koreans, whereas items such as *trellis* and *pelican*, which are meant to be relatively easier items, would be unfamiliar to Korean individuals and would in turn be more difficult.

The cultural specificity of certain BNT items is further demonstrated by the fact that individuals from different populations consistently produce alternate words or synonyms for certain items. Commonly produced synonyms among ethnic minorities in the US, particularly African-Americans, include *tommy walkers*, *tomwalkers*, or *walking sticks* for *stilts, falseface* for *mask*, and *harp*, *mouth organ*, or *French harp* for *harmonica* (Azrin et al., 1996; Strauss et al., 2006). In Australia, these alternate words or synonyms include *mouth organ* for *harmonica*, *squid* for *octopus*, *concertina* for *accordion*, and *lock* for *latch* (Cruice et al., 2000). In Canadian English, they include *dice* for *dominoes*, *walking stick* for *stilts*, *blues harp* for *harmonica*, *squeeze box* for *accordion* and *toadstool* for *mushroom* (Tombaugh & Hubley, 1997). Tombaugh and Hubley (1997) suggest following such responses with a probe such as, 'What is another name for this?'. However, if clinicians follow standard administration and scoring procedure, such synonyms would be considered incorrect, even though they may reflect regional or cultural variations rather than paraphasic errors (Azrin et al., 1996).

Thus, BNT performance is sensitive to word frequency and familiarity differences between cultures and languages. Such frequency and familiarity differences make it difficult for clinicians to determine whether the deficits seen on the BNT are due to real cognitive shortfalls, or to cultural-linguistic differences, or to some combination of the two.

Bilingualism. A further language variable that appears to influence performance on tests of visual confrontation naming is bilingualism. Studies suggest that bilinguals tend to perform more poorly than English monolinguals on the BNT (Kohnert, Hernandez, & Bates, 1998; Roberts, Garcia, Desrochers, & Hernandez, 2002); that they produce a different, more varied response pattern to BNT items (Kohnert et al., 1998; Roberts et al., 2002); that they may know the correct response in one language but not in the other (Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007); and that they benefit variably from a dual-language approach to administering and scoring confrontation naming tests (Gollan et al., 2007; Kohnert et al., 1998). Most studies that have investigated this issue have used Spanish-English bilinguals living in the US; however, Roberts et al. (2002) included a group of Canadian French-English bilinguals in their study and found that they also performed more poorly English monolinguals.

Commentary on these results has pointed to cultural, linguistic, and experiential factors (Kohnert et al., 1998). Some words are used more frequently or acquired earlier in English than in other languages, which means they are more familiar, and therefore easier to name, for English-speakers than speakers of other languages (de la Plata et al., 2009; Roberts et al., 2002). Due to this differing word salience between languages, simply translating a word from one language to another may result in the test not measuring the construct it was intended to measure in English by introducing language and cultural bias. One explanation offered for bilinguals' ability to produce the correct response in one language, but not the other, is the separation of work and home. Bilinguals may use different languages in each setting (e.g., Spanish at home and English at work), and may therefore be familiar with certain items only in the language used in the setting in which they are encountered (Bialystok & Craik, 2007; Gollan et al., 2007). That certain items are familiar only in one cultural or linguistic milieu may result in inaccurate BNT results for bilingual older adults.

Clearly, the generalizability of the above results regarding the effect of bilingualism on BNT performance is limited as the results may be specific to English-Spanish bilinguals. These results do caution, however, against the uncritical use of English-language norms for assessing the performance of individuals who are bilingual, even if they are proficient in

English. Clinicians need to consider patients' language profiles in assessing their performance on neuropsychological tests.

In summary, average BNT performance varies substantially across populations of different languages and cultures. This variation has called into question the appropriateness of using the BNT and its standardized norms with populations other than English-speaking North Americans, and has led to calls for representative culture- and language-specific normative data and suggestions that the test should be modified to replace problematic items when it is used in settings outside of the US. Following these suggestions, researchers and clinicians across the world have developed a number of culturally modified versions of the test.

Adaptation of the BNT for use with Diverse Cultures and Languages

There is a large body of literature that revolves around adaptations of the test for use in different cultures and languages. The enduring popularity and utility of the BNT is suggested by the fact that it has been translated and/or modified for use within many different cultures and languages, including Chinese (Cheung, Cheung, & Chan, 2004); Dutch (Mariën et al., 1998); Finnish (Laine et al., 1993); French-Canadian (Roberts & Doucet, 2011); Greek (Patricacou, Psallida, Pring, & Dipper, 2007; Simos et al., 2011); Hebrew (Kavé, 2005); Korean (Kim & Na, 1999); Portuguese (Miotto et al., 2010); Spanish in Argentina (Allegri et al., 1997; Serrano et al., 2001), in Spain (Peña-Casanova et al., 2009), and in the US (Pontón et al., 1992); and Swedish (Tallberg, 2005).

Researchers have adopted different approaches to the translation and cultural adaptation of the BNT, ranging from straightforward translations to extensive modifications. When modifying the test for the Swedish population, Tallberg (2005) translated the instrument and changed the sequence of items but did not find it necessary replace any items due to cultural bias. In order to deal with the cultural bias inherent in certain BNT items, others have replaced items from the original test with more culturally appropriate items, usually from the same semantic category (e.g., some have replaced *pretzel* with another type of food) and of the same difficulty and frequency. In their modification of the BNT for use with the Greek population, Patricacou et al. (2007) identified four problematic items (*pretzel, doorknocker, stethoscope,* and *scroll*); these items were replaced with traditionally Greek items from the same semantic categories (a kind of cake, mailbox, a blood pressure instrument, and an ancient Greek column). In creating an adapted BNT for Brazilian Portuguese-speakers, Miotto et al. (2010) replaced 20 items from the original 60-item BNT

with alternate items sourced from Brazilian books, newspapers, and television shows. A representative sample of 739 Brazilians aged 6-77 years scored significantly better on the adapted version than on the original BNT. Interestingly, when two items with low word-frequency in Australia were replaced with alternative items (*platypus* replaced *beaver* and *pizza* replaced *pretzel*), one study found that the change resulted in significantly improved scores among Australian elderly (Cruice et al., 2000), whereas another found that the alternative items made no difference to test scores (Worrall et al., 1995).

Other adaptions required much broader changes. In their development of a Korean version of the BNT (K-BNT), Kim and Na (1999) included only 10 items from the original BNT due to extensive cultural differences. They selected their items from an initial list of 175 items, sourced from dictionaries, textbooks, and print media. This list was cut to 70 items based on rigorous selection criteria and on the performance of a group of normal controls. A further 10 items were discarded based on the performance of a second group of controls and the same selection criteria. After rearranging the items in ascending order of difficulty, the authors found that word-frequency differed between Korean and English for the 10 original items included in the K-BNT. For example, the item *funnel*, which is item 49 in the original BNT, is item 10 in the K-BNT, showing it is relatively easier for this population than that for which the test was originally created. In creating a short form for use in China, Cheung et al. (2004) selected 30 items from the 60-item test, based on their cultural relevance, keeping the items in the same order as they appeared in the original test. However, in Cantonese, a logographic language, the names of objects usually contain one sound, which means that it is not possible to administer phonemic cues. The authors had to modify the test by replacing the phonemic cues with a multiple-choice task with two distractors (one semantically related to the target word and one perceptually related to it).

In summary, there is consistent evidence that the BNT needs to be modified to some degree if it is to be used outside of North America or with ethnic minorities inside the US. These attempts to modify the BNT are symptomatic of the problems being encountered with the international expansion of the field of neuropsychology and the resulting need for neuropsychological measures to be reliable and valid even when used outside of the country in which they were developed.

Current Issues in Cross-cultural Neuropsychology in South Africa

Neuropsychologists in lower middle income countries (LMICs), such as South Africa, tend to use standard neuropsychological tests developed in the global north in their

assessments of patients with cognitive or neurological impairment. However, such tests and their norms may have limited applicability in a multilingual and multicultural context like South Africa, overestimating an individual's level of impairment due to cultural and related biases.

South Africa has an extremely diverse population. There are 11 official languages (viz., Afrikaans, English, isiNdebele, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, SiSwati, Tshivenda, and Xitsonga), which are spoken with varying frequency across the country's nine provinces and different population groups (viz., Black African, Colored, Indian or Asian, and White). There is also significant diversity in the country with regards to education and socioeconomic circumstances, as a legacy of the country's political history. Just over one quarter of the population over 20 years of age (28.4%) have completed 12 years of schooling but roughly one tenth (8.6%) have received no schooling (Statistics South Africa, 2012). There are low levels of literacy and education among Black African adults whereas White adults have the highest level of education compared to other population groups (Posel, 2011; Statistics South Africa, 2012). Nearly half of the population lives in poverty, with disproportionately more Black African-headed households, female-headed households, less educated individuals, and individuals living in rural areas or certain provinces (viz., Kwa-Zulu Natal, Limpopo, and Eastern Cape) falling into this category than other groups (Armstrong, Lekezwa, & Siebrits, 2008).

A number of studies have highlighted the problems associated with using standard neuropsychological tests and their norms across these diverse ethnic, language, education, and socioeconomic groups in South Africa (Grieve & Cave, 2009; Owen, 1992; Robbins et al., 2013; Skuy, Schutte, Fridjhon, & O'Carroll, 2001). Without locally-appropriate normative data, or new or adapted culture-fair tests, clinicians working in South Africa may struggle to assess their patients fairly and accurately (Barratt, Khoza-Shangase, & Msimang, 2012; Nell, 2000; Shuttleworth-Edwards, Kemp et al., 2004). As Kalula et al. (2010) note, one of the central challenges facing clinical diagnosis in South Africa is that many of the tests in neuropsychological batteries may be inappropriate due to the low levels of education that are typical of patients seen at public clinics and health-care institutions in South Africa and the fact that many measures have not been validated for this population. Despite such findings, few neuropsychological tests have been developed, adapted, validated, or normed for South Africa's diverse population.

Those researchers who have approached the issue of cross-cultural neuropsychological testing have done so in at least two ways. They either use tests that are

cross-culturally fair, which may mean having to create new tests or adapt existing tests, or they create locally appropriate normative data for the standard versions of existing tests (Nell, 2000). Many South African researchers prefer re-norming existing neuropsychological tests for local populations, stratifying by core moderator variables, as they view this approach as more feasible and efficient than developing new, culture-fair tests from scratch (Shuttleworth-Jordan, 1996). Whichever option is followed, researchers in South Africa must acknowledge that performance on neuropsychological tests may be moderated by sociodemographic variables, including language, socio-economic status (SES), and quality of education (Grieve & Cave, 2009; Nell, 2000; Shuttleworth-Edwards, Kemp et al., 2004). Unfortunately, the number of published research studies presenting such data on neuropsychological tests in general, and on the BNT in particular, for the South African population is woefully inadequate. Hence, interpretation of test performance in this country is challenging.

In addition, although the provision of population-representative group norms can mitigate the impact of such influences on test performance and therefore improve the diagnostic accuracy of neuropsychological tests, this approach is not without problems. First, group norms do not address the underlying psychometric and sociocultural properties that explain why such between-group differences exist (Pedraza, Graff-Radford et al., 2009). Second, the lack of cultural equivalence in the test is not addressed by the provision of separate norms, i.e., it does not improve the construct validity of the test (Manly, 2005). The collection of representative normative data is one step in addressing the cross-cultural and cross-linguistic problem in neuropsychological testing, but of more significance is investigating the validity of such measures in these population groups (Kohnert et al., 1998).

Research on the BNT in South Africa. Although some studies have validated modified BNTs for use within specific cultures and languages, until recently there was no research available on the use of BNT in South Africa. Those studies that have investigated the performance of South Africans on the BNT generally confirm the findings emerging from other non North-American countries, of cultural and linguistic bias.

In an unpublished Master's thesis, Mendonça (2010) investigated the performance of 116 English-, Zulu-, and Sotho-speaking South African university students (aged 18-21 years) on the 60-item BNT, administered in English. The author found that the South African sample as a whole scored significantly more poorly than Canadian norms on the 60-item

BNT, with 40 items identified as being inappropriate[1] for use in South Africa. English-speaking students scored significantly better than Zulu- and Sotho-speaking students, but, on average, the latter two groups scored similarly. Mendonça reported that monolingual students scored significantly higher than bilingual students. Although this finding is consistent with other investigations of bilingualism on test performance (see, e.g., Kohnert et al., 1998), the study had various methodological weaknesses. For instance, the monolingual group was comprised of English-speaking participants only, and they had already been shown to perform better than both Zulu and Sotho participants.

In the only published study investigating a South African adaptation of the BNT, Mosdell, Balchin, and Ameen (2010) created the Groote Schuur Naming Test (GSNT) by adapting the B. Williams et al. (1989) odd-numbered 30-item test. Each item in the original test was replaced with one thought to be more specific and familiar to South African cultures, predominantly through consultation with cultural and language experts. The authors translated both the original and modified tests from English into Afrikaans and isiXhosa. The two tests were then administered to a sample of 30 neurologically intact orthopedic patients in their home language, with an equal division of English, Afrikaans, and isiXhosa speakers.

Results suggested that the BNT contained language and cultural bias. English and Afrikaans participants scored similarly to each other on the B. Williams et al. (1989) 30-item BNT; however, there was a significant difference between English speakers and isiXhosa speakers, with the latter performing significantly more poorly and below US normative standards for AD patients. The authors identified the items *trellis, asparagus, pyramid, hammock, sphinx, unicorn, pelican,* and *beaver* as the most problematic. Unsurprisingly, all of the participants were able to name significantly more items correctly on the GSNT than on the BNT. Even though South African-specific replacement items were selected for the GSNT, the authors also identified a set of problematic items on this test (for instance, the items *mug*, *dragonfly*, and *hippopotamus* were misnamed frequently, particularly by the isiXhosa-speaking participants).

Summary and Rationale for the Present Study

In South Africa, there are few locally developed, modified, validated, or normed neuropsychological tests. The cognitive tests used most widely by clinicians and researchers in this country were developed and normed in the US. However, such cognitive tests and

---

[1]An item was defined as inappropriate if there was a statistically significant difference in the proportion of correct responses between the South African and Canadian samples.

their accompanying norms cannot simply be appropriated from one culture or language and applied directly to another. When used uncritically in LMICs such as South Africa, where much of the population is culturally and linguistically different from the population in the country where the test originated, and where lower levels of education are common, test results have the potential to misinform and to lead to distorted diagnostic accuracy (Anderson, 2001; Nell, 2000). There is a pressing need amongst South African clinicians for cognitive tests appropriate for assessment in the local context, and for relevant normative data, which is essential to ensure the diagnostic utility of such tests (Barratt et al., 2012; Shuttleworth-Jordan, 1996). The scope for research in this field is therefore wide.

The BNT is already used in South African clinical settings (Kalula et al., 2010) and research studies (Nield, 2007), despite the fact that such use may be inappropriate because of the issues described above. With the development of neuropsychology as a profession in South African - and therefore with increasing numbers of neuropsychologists being trained and working in the field in this country - the need for valid and reliable tests for use with the South African population is growing. In addition, there is growing overseas interest in South African-based research into conditions such as HIV and fetal alcohol syndrome (Ferrett, Carey, Thomas, Tapert, & Fein, 2010; Robbins et al., 2013). Oftentimes, the investment of grant money into such research endeavors is contingent upon the use of psychometrically sound instruments that are also well known to international audiences.

Although there is extensive research on the BNT elsewhere in the world, there is a dearth of research on the test in South Africa. Thus, researchers from the University of Cape Town (UCT) and the University of Stellenbosch (US) embarked on a project seeking to modify, translate, and provide appropriate normative data for the BNT (and other commonly used neuropsychological tests) for English, Afrikaans, and isiXhosa-speaking individuals in the Western Cape. This project led to the development of a modified short form, the Boston Naming Test-South Africa short form (BNT-SA-SF), which includes 15 items judged by a forum of practicing clinical neuropsychologists as being more culturally appropriate for the South African population than those on the most popular 15-item short form (the Mack SF-4; Mack et al., 1992).

Many of the modified forms of the BNT, including Mosdell et al.'s (2010) naming test, have substantially changed the items from the original test. Although these modifications seek to make the test appropriate for local populations, they make comparisons between studies, for cross-cultural or cross-linguistic purposes, difficult. In addition, 60-item or 30-item modifications are not always practical for use in hospital settings or in research test

batteries, particularly if they are to be used in assessing people with dementia who have a short attention span. The value of the short form described here lies in the fact that (a) it only includes 15 items, and (b) the items are drawn from the original test. The latter means that the test is readily available to clinicians or researchers who already possess the original, and is advantageous for comparative purposes.

Despite the utility of such a short form in clinical and research settings, the psychometric properties of the test have not been investigated adequately. If it is to be utilized by clinicians in the Western Cape, and in South Africa more broadly, such information is vital. Thus, this research attempts to assess the psychometric properties of the BNT-SA-SF in a clinical setting. As the BNT is most frequently used in the assessment of patients with dementia, and has shown particular discriminative utility in identifying the naming deficit present in AD, the present study aims to validate the use of the BNT-SA-SF as a screening measure to help identify AD from normal aging and other types of dementia. The results of this study should aid in the accurate interpretation of BNT-SA-SF performance in older adult populations, particularly those in the Western Cape and in those with possible or probable AD.

Specific Aims and Hypotheses

The broad aim of this study was to investigate the psychometric properties of the BNT-SA-SF, focusing on the diagnostic validity of the test. Evaluating the validity of a test in the context in which it is to be used is an important part of test construction. I chose to evaluate the specific clinical application of using the test to classify individuals dichotomously into normal versus impaired naming performance, examining the performance of patients from a Cape Town Memory Clinic and matched controls. To assess the diagnostic validity of the BNT-SA-SF, I addressed a number of main questions. Some of these questions have specific hypotheses or predictions attached to them, but other questions are purely exploratory. The questions are as follows:

(1) What is the internal consistency reliability of the BNT-SA-SF?

(2) What is the discriminative capacity of the BNT-SA-SF? That is to say, how well can it distinguish between (a) patients with AD and controls, (b) patients with AD and patients with dementia diagnoses other than AD, and (c) patients with dementia diagnoses other than AD and cognitively intact controls?

    a. With regard to BNT-SA-SF total score, my specific hypotheses were that (i) patients with AD will perform significantly more poorly than cognitively

intact controls, (ii) patients with AD will perform significantly more poorly than patients with dementia diagnoses other than AD, and (iii) patients with dementia diagnoses other than AD will perform significantly more poorly than cognitively intact controls.

    b. With regard to the sensitivity and specificity of the test in the differential diagnosis of dementia, I hypothesized that the test will have the strongest diagnostic accuracy in distinguishing patients with AD from cognitively intact controls, compared to (i) patients with AD from patients with dementia diagnoses other than AD, and (ii) patients with dementia diagnoses other than AD from cognitively intact controls.

(3) Do age, education, sex, race, language, and SES affect BNT-SA-SF performance? Specifically, are there real between-group differences in BNT-SA-SF performance, or do sociodemographic variables influence performance on the test? Based on previous literature, I predicted that BNT-SA-SF performance will

    a. decline with increasing age

    b. improve with increasing levels of education

Furthermore, I predicted that (c) participants who speak a language other than English most frequently will perform more poorly than participants who speak English most frequently, and (d) participants of lower SES will perform more poorly than patients of higher SES.

(4) Is there a relationship between BNT-SA-SF performance and dementia severity in AD? I hypothesized that BNT-SA-SF performance would decline with increasing dementia severity.

(5) What is the pattern of performance across the 15 test items? Further, what is the relationship between item difficulty and item placement in the test? Are there any items that appear problematic? With regard to item performance, I hypothesized that item difficulty would increase in linear fashion from item 1 through to item 15.

Methods

Design and Setting

This study used a quasi-experimental non-equivalent groups design in accordance with nonrandomized selection criteria. I compared patients, based on nonrandomly selected archival records, from a Cape Town Memory Clinic with healthy, community-dwelling control participants on a modified BNT short form. There was a single administration of all measures to newly recruited participants

Testing of the patient sample took place at the Memory Clinic, where a qualified neuropsychologist or a well-trained and experienced neuropsychology Master's student administered the tests. Because control participants were drawn from various sources, testing took place in a number of settings, including the Memory Clinic and participants' places of residence. A Master's student, trained and experienced in administering the neuropsychological tests used in this study, administered all the measures to control participants.

Participants

Patient group. The patient group was comprised of individuals who were referred to the University of Cape Town/Groote Schuur Hospital Memory Clinic, a program of the Albertina and Walter Sisulu Institute of Ageing in Africa (IAA). The Memory Clinic is a weekly, half-day, outpatient clinic held at Groote Schuur Hospital's Department of Psychiatry. The clinic, established in 1999, serves patients from Cape Town and the broader Western Cape region. The health professionals who work at the clinic seek to diagnose and to provide referrals for treatment options for older adults with dementia-related cognitive decline. Patients who present at the Clinic have been referred by a medical practitioner because of noticeable memory loss and/or other behavioral and psychological symptoms of dementia (Kalula et al., 2010). Each patient is assessed in a single consultation by a multidisciplinary team, including a geriatrician, a psychiatrist, and a neuropsychologist. The clinical team make diagnoses based on a structured clinical interview, thorough neuropsychological assessment, and physical examination, according to standard criteria and by team consensus. The neuropsychological assessment battery includes a number of commonly used neuropsychological tests, as well as a depression scale, activities of daily living scale and caregiver distress scale. In other words, diagnosis is based on a thorough, comprehensive assessment, and does not rely solely on the BNT-SA-SF or MMSE, the two tests referred to in this study.

I examined Memory Clinic archival data for all patients admitted from January 2010[2] to September 2012. Patients were included in the final sample if item-level data were available for their BNT-SA-SF administration[3]. Data from that sample of patients ($n = 153$) was then divided as follows. First, data from those patients with insufficient diagnostic and/or demographic information ($n = 45$) were excluded from the sample. The latter information is important for matching purposes and for investigating the relative influence of key demographic variables on BNT performance. Second, data from those patients not diagnosed with dementia ($n = 35$) were excluded from the sample (a comorbid psychiatric diagnosis (e.g., major depressive disorder alongside AD) did not warrant exclusion). Third, data from patients who indicated during the clinical interview that they did not speak English ($n = 4$) were excluded from the sample. Many of the tests included in the Memory Clinic neuropsychological test battery are only available in English (because they were developed in the United States or the United Kingdom) and are administered in that language if the patient indicates that they are fluent in English (because Xhosa- and Afrikaans-speaking neuropsychologists are not always present at the Clinic). Although Memory Clinic clinicians are careful to consider this factor when assessing and diagnosing individual patients, results for these individuals are generally confounded by the impact of language difficulties on test performance. Afrikaans, isiXhosa, and English are the three main language groups in the Western Cape, with 49.7%, 24.7%, and 20.2% of the population in this province speaking each language as their home language respectively (Statistics South Africa, 2012). However, as English is the primary language used in education in the province and nationally, many people are fluent in English even if this is not their home language.

Fourth, data from the remaining patients ($n = 69$) were divided into two groups, according to diagnosis. The Alzheimer's disease (AD) group ($n = 46$) was comprised of patients diagnosed with possible or probable AD alone or in combination with VaD. It would be difficult to construct a pure AD group due to the overlap between these two diagnoses and the fact that both lead to naming deficits (de la Torre, 2004; Miller et al., 2010). The other dementia (OD) group ($n = 23$) was more heterogeneous and was comprised of patients diagnosed with dementia not of the Alzheimer's or mixed type. The most common diagnosis at the Memory Clinic is dementia, with 74% of patients being diagnosed as such, and 59% of these being diagnosed with the predominant subtypes, AD or mixed AD/VaD (Kalula et al.,

---

[2] Prior to January 2010, another BNT short form, the Mack SF-4 (Mack et al., 1992) was administered at the clinic as the confrontation naming test in the neuropsychological test battery.

[3] Not all patients who present to the clinic are administered a confrontation naming test due to their level of impairment, time constraints, or language difficulties.

2010). The other most commonly diagnosed dementia at the clinic is VaD, but cases of Lewy Body dementia (LBD), Parkinson's disease with dementia (PDD), frontotemporal dementia (FTD), and alcohol-related dementia, among others, are also seen routinely.

Figure 1 is a schematic explaining the manner in which the final sample was created. In the AD group, 34 participants (73.9%) were diagnosed with possible or probable AD and 12 (26.1%) were diagnosed with mixed AD/VaD. In the OD group, 15 participants were diagnosed with VaD (65.2%) and 8 (34.8%) were diagnosed with other types of dementia, namely alcohol-related dementia ($n = 1$), FTD ($n = 1$), LBD ($n = 3$), PDD ($n = 1$), and dementia undetermined ($n = 2$).

Control group. Healthy, community-dwelling older adults formed the control group. These participants were recruited in a number of ways. First, I approached individuals accompanying Memory Clinic patients to their appointments and asked them to participate ($n= 25$). A spouse, family member, or carer who is cognitively intact and who is therefore an appropriate control participant usually accompanies patients.

Second, I recruited participants from a housing complex in Cape Town ($n = 26$). This housing complex is run by a social housing company that provides rental accommodation and basic social services, primarily to older adults, in various residential suburbs across Cape Town. I chose this specific complex on the basis that it was thought most likely to contain a similar demographic of older adults to those who present at the Memory Clinic.

The use of three groups (AD, OD, and controls) is similar to the procedure used by previous studies validating new BNT short forms (e.g., Graves et al., 2004; B. Williams et al., 1989). Specifically, Graves et al. (2004) used an AD/VaD group, comprised of patients diagnosed with AD or mixed AD/VaD, a group of patients with other diagnoses, comprised of other medical disorders and dementia types other than AD or AD/VaD, and patients diagnosed as normal, in the validation of 10 existing short forms and 3 newly created short forms.

Inclusion and exclusion criteria. Eligibility criteria were applied to ensure results were not confounded by extraneous variables and to guarantee a representative, healthy sample that was of a similar demographic composition to the patient sample. The inclusion criteria specified that control participants needed to reside in the Cape Town metropolitan region, be older than 50 years of age, and speak English.
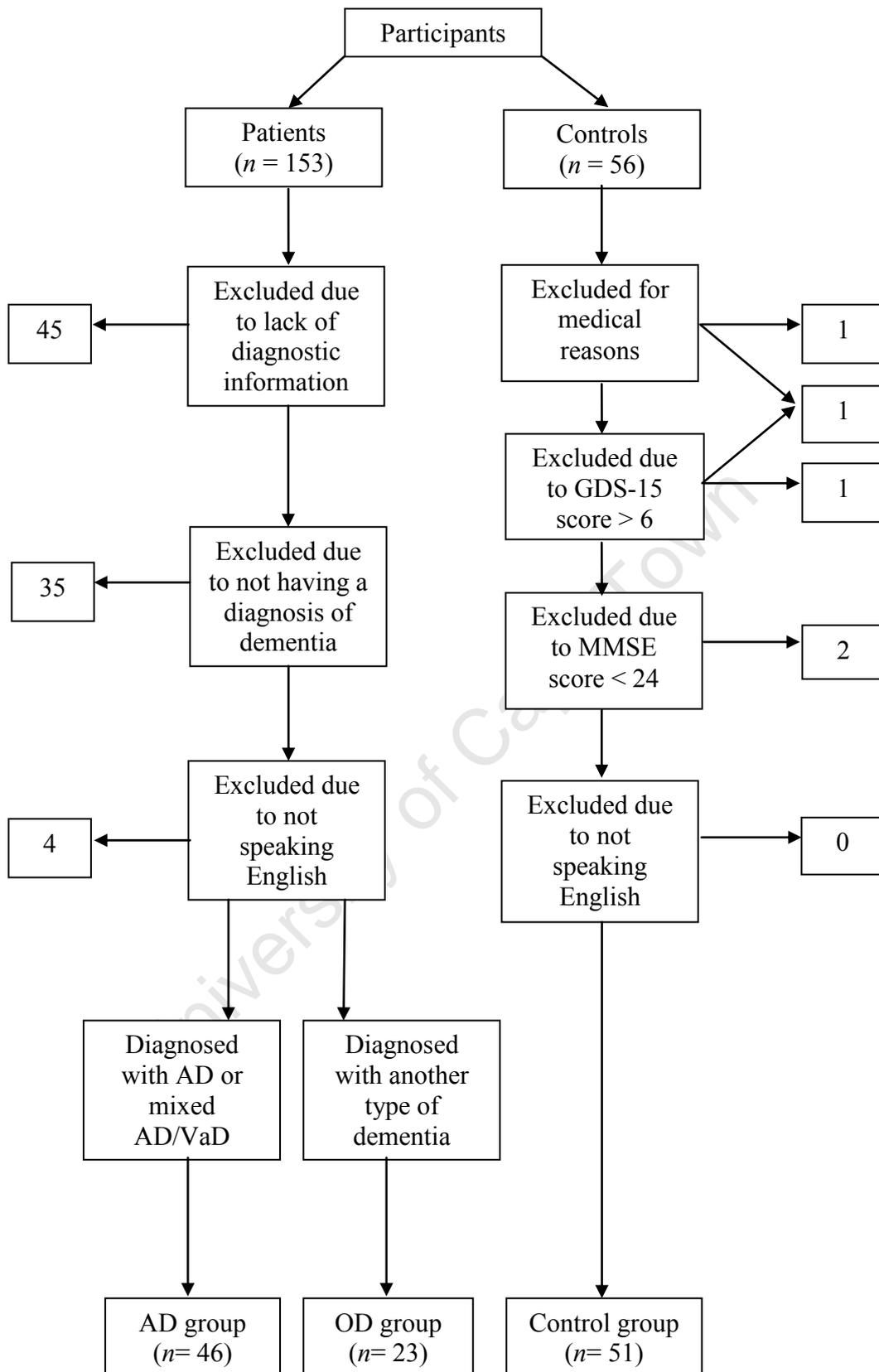
Figure 1. Schematic explaining how the final patient and control samples were produced.

Exclusion criteria were (a) a score < 24 on the MMSE, which is the most widely used cut-off score on this test indicating cognitive impairment; (b) having received a psychometric evaluation in the past 12 months, to rule out the possible effects of practice on test performance; (c) a score > 6 on the 15-item Geriatric Depression Scale (GDS-15; Sheikh & Yesavage, 1986), which indicates clinically relevant depressive symptoms; and (d) residing in an assisted-living facility, to ensure a reasonable degree of independence in daily functioning.

To ensure that participants were physically healthy, I obtained an extensive medical history via self-report questionnaire (see Appendix A). The following criteria warranted exclusion from the study: current use of psychotropic medication, and/or a history of psychiatric diagnosis; a head injury that resulted in a loss of consciousness for more than 5 minutes; seizure disorders; a serious medical illness that may affect neuropsychological functioning, such as meningitis, multiple sclerosis, or encephalitis; and learning, language, speech, or other educational difficulties. The above criteria were used because they influence performance on neuropsychological tests, including the BNT (Lezak et al., 2004; Strauss et al., 2006).

Initially, I recruited 56 controls. From that sample, one potential participant (a male aged 74) was excluded for having had a head injury that resulted in continuing seizures, one (a female aged 51) was excluded for having a history of epilepsy and a GDS-15 score > 6, one (a female aged 52) was excluded for having a GDS-15 score > 6, and two (a male aged 83, and a female aged 65) were excluded due to MMSE scores < 24. No control participants were excluded due to language criteria, as all indicated they were fluent in English. Figure 1 also shows a schematic explaining the manner in which the control group was arrived at.

Materials

Demographic form. All control participants were required to complete a sociodemographic questionnaire (see Appendix A). The questionnaire gathered information about the participant's race, sex, age, educational background, and socioeconomic circumstances. I used this information to match the control and patient samples, and, during data analysis, to investigate the impact of demographic variables on test performance. The questionnaire was also designed to gather information about the participant's medical, psychiatric, psychometric, and scholastic history. I used this information to ascertain whether the participant met the eligibility criteria described above.

Memory Clinic patients were not administered these forms, but the clinicians routinely gather similar information during the structured interview portion of the appointment. That information is therefore contained in each patient's Memory Clinic file.

Screening measures. A set of standardized instruments was used to further screen the potential control participants. These instruments included measures of general cognitive functioning (MMSE; Folstein et al., 1975), depressive symptoms (GDS-15; Sheikh & Yesavage, 1986), and activities of daily living (modified Bristol Activities of Daily Living Scale (BADLS); Bucks, Ashworth, Wilcock, & Siegfried, 1996). These measures were included to ensure a healthy, non-impaired sample, and because the same or similar measures are used at the Memory Clinic.

Mini-Mental State Examination. This is the most widely used screening measure to assess general cognitive functioning in older populations. The test has 19 individual questions covering the domains of orientation, registration, attention and calculation, recall, and language (naming, repetition, verbal and written comprehension). It is scored out of 30, with lower scores indicating greater impairment. It has been shown to be a reliable and valid measure in detecting cognitive impairment in elderly populations (Mitchell, 2009; Tombaugh & McIntyre, 1992), and it is one of the key components of the Memory Clinic neuropsychological test battery. In memory clinic settings it has a sensitivity of 79.8% and a specificity of 81.3% in detecting dementia and a sensitivity of 85.1% and a specificity of 85.5% in detecting dementia in community settings (Mitchell, 2009).

The test is used extensively in international and South African clinical and research settings. Recent published studies have reported data on the performance of the test with English- and Afrikaans-speaking older adults, similar to the demographic of the sample in this study (Roos et al., 2010; van Schalkwyk, Botha, & Seedat, 2012). These local studies have, similar to previously published studies (see Tombaugh & McIntyre, 1992, for a comprehensive review), observed a positive relationship between MMSE score and level of education. Hence, when interpreting MMSE performance it is important to account for education effects, particularly in populations where the average level of education is low (Grigoletto, Zappalà, Anderson, & Lebowitz, 1999).

Generally, a score of < 24 is accepted as indicating cognitive impairment in individuals with 8 or more years of education (Lancu & Olmer, 2006; Tombaugh & McIntyre, 1992). Thus, to be on the safe side for the purposes of this study and to ensure the integrity of the control sample, I used a score of 24 as an eligibility cut-off for control participants. In the current administration, both the attention and calculation tasks were

included, with the higher of the two scores being used in calculating the total score. This procedure follows the recommendations of Tombaugh and McIntyre (1992), and is the same procedure followed by others (Ashford, Kolm, Colliver, Bekian, & Hsu, 1989; Hawkins, Cromer, Piotrowski, & Pearlson, 2011). The lack of consensus regarding the administration of these two tasks is an acknowledged weakness of the test, however, as the serial 7s subtraction task appears to be significantly more difficult than the "world" backward task (Hawkins et al., 2011).

Geriatric Depression Scale. Depressive disorders are common in the elderly. For instance, results of the South African Stress and Health (SASH) study, a nationally representative household survey, found lifetime prevalence rates of 10.0% in those 50-64 years and 6.5% in those 65 and older (Herman et al., 2009). Hence, I included a depression scale in the screening of control participants.

The GDS-15 (Sheikh & Yesavage, 1986) is a self-report scale consisting of 15 questions, each with a yes/no response format. Respondents are asked to choose the best answer for how they have felt over the past week, with 10 of the items indicating depression when answered positively and the other 5 indicating depression when answered negatively. Higher scores indicate more severe depressive symptoms.

The original test (Yesavage et al., 1983), developed specifically for depression screening in elderly subjects, had 30 items but the shorter GDS-15 is probably the most popular version of the test, taking only 5-7 minutes to administer. The short form correlates strongly with the full version (Burke, Roccaforte, & Wengel, 1991; Sheikh & Yesavage, 1986), and with other gold standard depression assessments (Herrmann et al., 1996; Marc, Raue, & Bruce, 2008). It has good psychometric properties across different populations (Pedraza, Dotson, Willis, Graff-Radford, & Lucas, 2009; Wancata, Alexandrowicz, Marquart, Weiss, & Friedrich, 2006). Although originally designed and validated as a self-administered scale, the psychometric properties of the test remain strong even when administered orally (D'ath, Katona, Mullan, Evans, & Katona, 1994; Herrmann et al., 1996).

For the purposes of this study, a cut-off score of 6 warranted exclusion; 5 or 6 are the most widely reported cut-off scores yielding optimum levels of sensitivity and specificity (Wancata et al., 2006).

Bristol Activities of Daily Living Scale. Activities of daily living (ADL) scales assess an individual's daily self-care activities as a measure of their functional state, and are most often used in assessing the elderly. Impaired ADL is one of the key features of AD and other types of dementia according to the *Diagnostic and Statistical Manual of Mental Disorders*

(5th ed.; *DSM-5*; American Psychiatric Association, 2013) and the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association working group (NINCDS-ADRDA; McKhann et al., 1984) criteria. The assessment of ADL is therefore useful in early screening for dementia, especially AD (Gauthier, Gélinas, & Gauthier, 1997).

The BADLS (Bucks et al., 1996) is a carer-rated questionnaire, developed specifically for use in screening possibly demented individuals within clinical research, hospital, or community clinic settings. The version of the BADLS used in this study, which includes 17 items from the original 20-item scale, was modified for use at the Memory Clinic as part of the standard assessment protocol to assess basic (e.g., preparing food) and instrumental (e.g., managing finances) ADLs. It has a minimum score of 0 (indicating an individual who is totally independent in performing ADLs) and a maximum score of 51 (totally dependent).

The BADLS is a reliable and valid measure (Bucks et al., 1996; Sikkes, De Lange-de Klerk, Pijnenburg, & Scheltens, 2009). It correlates well with the MMSE, and is sensitive to a range of ADL performance in individuals with different levels of functioning, from those that are completely independent to those that are completely dependent (Bucks et al., 1996).

In this study, the modified BADLS was used as a measure of control participants' ability to live independently and with a high level of functioning in their daily lives. As noted earlier, control participants were screened according to the demographic questionnaire and were specifically sampled to ensure that they were likely to be independent and healthy, with a high level of everyday functioning, by excluding participants if they lived in assisted-living facilities. Thus, the modified BADLS was simply used as an additional measure to make certain that this was the case. No cut-off score was used, and participants were assessed on a case-by-case basis. All control participants except for one, who scored 6 due to a recent injury temporarily affecting her mobility, scored < 2.

Although the original BADLS was developed as a caregiver-rated instrument, the modified BADLS was administered with the control participant themselves as many lived alone, which meant that the administration of such a scale with a caregiver was not feasible. For this reason, there were some concerns that participants would overrate their abilities; however, due to the other measures in place to effectively screen control participants, I did not consider this a significant concern.

BNT-SA-SF. The BNT-SA-SF (see Appendix B) comprises 15 items drawn from the pool of 60 items included in the long-form administration of the BNT-2. The aim of creating this modified version of the test was to decrease cultural bias and to ensure equivalent

performance across the three languages that predominate in the Western Cape province of South Africa (English, Afrikaans, and Xhosa). To create this short form, researchers from our laboratory divided the 60 BNT items into 15 sets of four items of equivalent difficulty. For example, items 1, 2, 3, and 4 formed a pool of items; similarly, items 5, 6, 7, and 8 formed a pool. In this way, each of the 60 items was assigned to 1 of 15 pools.

From each of these pools, a panel of 15 South African neuropsychologists, all of whom use the BNT in their clinical practice, selected one item. Thus, the item they chose from the first pool formed item one of the modified test, the item they chose from the second pool formed item two of the modified test, and so forth. This procedure attempted to ensure the items in the short form were of increasing difficulty in a similar manner to the original test, and is comparable to the procedure followed by Mack et al. (1992) in the creation of their short forms.

The neuropsychologists made their selection by rating each item in the 15 pools according to picture quality or ambiguity, cultural familiarity or appropriateness, rank (difficulty), and colloquial use. Hence, the item considered most appropriate in each pool was selected for the BNT-SA-SF. Finally, a team of linguists translated and backtranslated the test items and administration instructions for the modified test from English into the other two languages most widely spoken in the Western Cape (Afrikaans and isiXhosa).

Procedure

Patients. Participants in the patient group were tested as part of their neuropsychological assessment on presentation at the Memory Clinic. During the appointment, a team of professionals, including a medical doctor, a neuropsychologist, and a psychiatrist, examine the patient and come to a diagnosis via consensus. One facet of this examination is the administration of a standard screening battery of neuropsychological tests, which includes the MMSE and BNT-SA-SF.

I used demographic information about the patient that is routinely collected during the appointment and filled into dedicated Memory Clinic booklets. I obtained these files containing the patients' demographic details and test scores from the Institute of Aging in Africa at Groote Schuur Hospital, where they are stored.

Controls. As noted above, I recruited control participants in two ways. Individuals accompanying Memory Clinic patients to their appointments and who were of an appropriate age (i.e., older than 50 years) were informed about the study and offered the opportunity to participate. If they took that opportunity, they were tested individually in a separate, quiet

room while the patient was receiving his/her physical examination and being administered the neuropsychological test battery.

The procedure was slightly different for the recruitment of participants from the housing complex. First, I contacted the housing complex by sending a letter of introduction (see Appendix C) to the administration of the social housing company that managed this facility, as the organizational body had to grant permission. Following their approval, I contacted the facility and distributed flyers advertising the study (see Appendix D). I then held an information session for interested residents. That session allowed me to inform them about the study and to invite them to sign up for participation by providing their names and contact details. After I made appointments with the individuals who had signaled their interest, I tested each individually in a quiet room at his/her place of residence.

The remainder of the test procedure was the same for all control participants. I asked them to complete a consent form (see Appendix E) and the demographic questionnaire. Although Afrikaans translations of the consent form and demographic questionnaire were available, all participants indicated that they preferred to receive the forms in English. I then administered the MMSE, GDS-15, BADLS, and BNT-SA-SF, in that order. Clear instructions were given before each test was administered. After test administration was concluded, participants were fully debriefed and given the opportunity to ask questions or express opinions about the test procedure.

BNT-SA-SF administration. The BNT-SA-SF was administered and scored identically for all participants. These administration and scoring procedures followed standard BNT conventions. That is, if the item was correctly named spontaneously within 20 seconds, the examiner proceeded to the next item. Spontaneous self-corrections were also scored as correct. If an incorrect response or no response was given, the examiner followed one of two responses. If an incorrect response that indicated misperception or no response was given, the semantic cue was given. If the incorrect response was semantically similar, but not the correct work, e.g. 'chair' for 'bench', an eliciting cue such as, 'can you think of another name for it?' was given. This, although not in the instructions in the test stimulus booklet, is recommended by a number of researchers utilizing the test (e.g., Tombaugh & Hubley, 1997). If the participants still could not produce the correct response within the given 20 seconds, the examiner provided the phonemic cue. If the participant did not respond correctly after being given that cue, the examiner gave a phonemic cue. Regardless of whether the participant responded correctly after being given that cue, the examiner moved on to the next item. After all the items were administered in this way, the examiner returned

to all items for which a correct response was not produced after the phonemic cue. For each of those items, the examiner showed the picture to the participant again and read out four multiple-choice options from which the participant was asked to select the one that described the pictured object best.

The overall test score was calculated, according to standard BNT scoring, by adding correct responses made spontaneously or following a semantic cue. Hence, each participant received one score out of 15. The shortened name for rhinoceros, 'rhino', was also marked as correct.

One way in which the current administration differed from standard BNT administration protocols involved language. Because Memory Clinic patients who are fluent in English are usually administered the tests in English, irrespective of their home language, this same procedure was applied when testing controls. However, because the BNT-SA-SF has been translated into Afrikaans and Xhosa, correct answers given in either of these languages also contributed to the final score. This procedure ensured that the test was administered and scored in the same way to control participants as it is administered to patients at the Memory Clinic.

Ethical Considerations

Patients. As noted above, patients referred to the Memory Clinic are administered the MMSE and BNT-SA-SF as part of a neuropsychological test battery that forms one component of a routine multidisciplinary assessment of all new admissions. As part of giving their consent to receive clinical service delivery at the Memory Clinic, patients give their permission for clinic-affiliated researchers to use data derived from their assessment. The UCT Faculty of Health Science Research Ethics Committee approved this data collection procedure.

Each patient included in the final sample was assigned a Study ID number. Hence, no names or other identifying information were used in the data analysis and write-up.

Controls. Because the testing procedure for controls was relatively short and straightforward, without any deception, and devoid of any invasiveness, I anticipated no negative consequences. However, a number of ethical considerations must be noted. Consent was obtained from each participant before the demographic form was completed or any of the tests were administered. The examiner clearly explained to participants that they could withdraw from the study at any time and that there would be no negative consequences

should they wish to do so. Furthermore, the examiner guaranteed all participants the confidentiality of their personal details, such as those that appeared on the demographic form.

The examiner debriefed each participant after completion of the tests. They were given an opportunity to ask any questions that may have arisen during the testing experience. Participants were then thanked for participating in the study, and the examiner explained to them how to access the study's results when available. The Research Ethics Committee of the University of Cape Town's Department of Psychology granted approval to collect data from control participants.

Each control participant was given a Study ID number. Hence, no names or other identifying information were used in the data analysis and write-up.

Follow-up procedures. A further source of concern was that control participants might perform poorly on the screening measures and/or BNT-SA-SF, thus indicating possible depression or cognitive impairment. Although the demographic form was used to screen patients for neurological or cognitive problems, a history of psychiatric diagnoses, or use of psychotropic medication, I acknowledged that this screening might not identify impairment in all cases and thus put a number of procedures in place to deal with such circumstances.

Cognitive impairment. If participants scored below the most widely used cut-off score for cognitive impairment on the MMSE (i.e., < 24), the examiner gathered a short history from the participant. With the participant's consent, this history, their demographic details, and test scores were sent to the Memory Clinic to be reviewed by a member of the clinical team. If it was deemed necessary, the participant was referred for a clinic appointment. Following this procedure, two control participants (a male aged 83 and a female aged 65) were referred to the Memory Clinic.

Depression. If participants scored more than six on the GDS-15, the examiner also gathered a short history of their feelings of depression. I then contacted the participants who met this criterion at a later stage and gave them advice on how to access support services for depression. For those participants recruited from the social housing program retirement complex, consent was obtained to communicate their details to the social worker from the housing program's area office. That office provides various psychosocial and recreational services to residents from the complexes under its management. The social worker then followed up with these individuals to ensure that they obtained the appropriate support. Two potential control participants (both females aged 51 and 52 respectively) who scored > 6 on the GDS-15 were provided support in the above manner.

Data Management and Statistical Analysis

I analyzed the data using the SPSS software package, version 21.0. The threshold for statistical significance (α) was set at .05, and appropriate effect size estimates were calculated for each analysis. In the initial analytic step, I generated the descriptive statistics for all relevant variables, and examined the assumptions underlying the necessary statistical tests. For continuous variables, I used the relevant parametric tests if assumptions underlying those analyses were upheld; for categorical variables, I used chi-square tests if assumptions underlying such analyses were upheld. For data not meeting the necessary assumptions, I used alternative robust tests or equivalent non-parametric tests. There were several components to the analytic strategy that followed. Each step is outlined below.

Step 1: Sample characteristics. This step assessed the sociodemographic composition and MMSE performance of the patient and control groups. A series of between-group comparisons ensured that the AD, OD, and control groups were matched on key variables that, given the results from previous BNT normative studies (e.g., Tombaugh & Hubley, 1997; Zec et al., 2007a) and numerous South African studies on neuropsychological tests (e.g., Grieve & Van Eeden, 2010; Skuy et al., 2001), might have influenced test performance. I used Pearson's chi-square analyses or, if the data did not meet the assumptions underlying that analysis (where more than 20% of the cells had minimum expected cell counts < 5), Fisher's exact test, on categorical variables (viz., sex, race, level of education, SES, and language). I used ANOVA or Welch's F (B. Welch, 1951), which is robust when the assumption of homogeneity of variance has been violated, on continuous variables (viz., age and MMSE score).

Step 2: Internal consistency. I assessed the internal consistency reliability of the test by calculating Cronbach's alpha for the entire sample. Cronbach's alpha is used routinely as a measure of internal consistency in the assessment of new and existing tests' validity. For each item, I also examined the value of Cronbach's alpha if that item was deleted.

Step 3. Between-group differences in BNT-SA-SF performance. As the first step in assessing the diagnostic ability of the BNT-SA-SF, I used Welch's F to investigate the hypotheses that (a) patients with AD will perform significantly more poorly than cognitively intact controls, (b) patients with AD will perform significantly more poorly than patients with dementia diagnoses other than AD, and (c) patients with dementia diagnoses other than AD will perform significantly more poorly than cognitively intact controls. In other words, this analysis sought to determine whether BNT-SA-SF performance (i.e., total number of correct responses on the test) could distinguish the three groups.

Step 4. General linear model: Relationship between BNT-SA-SF performance, and sociodemographic variables. A theoretically-guided exploratory procedure, using a general linear model (GLM), described the influence of key sociodemographic variables and MMSE performance on the discriminability of the BNT-SA-SF. Specifically, I wanted to test whether any of the demographic variables that previous literature suggests might influence BNT performance could (either singularly or in interaction) account for a significant proportion of the variance in BNT-SA-SF score, even when taking group status (AD, OD, and control) into account.

As an initial modeling step, I entered age, level of education, sex, language, SES, race (as a proxy for quality of education), and group status, with two-way interactions, as predictor variables. I then removed non-significant variables one by one, starting with the most complex (e.g., 2-way interactions) and least significant, and worked iteratively toward a statistically significant model that explained the most variance in BNT-SA-SF score.

Step 5: Receiver operating characteristic (ROC) analyses. ROC curves, with corresponding area under the curve (AUC) values, examined the classification accuracy of the BNT-SA-SF for distinguishing (a) patients with AD from controls, (b) patients with AD from patients with dementia diagnoses other than AD, and (c) patients with dementia diagnoses other than AD from cognitively intact controls. I also used the ROC curves to identify the optimal cut scores and to evaluate the sensitivity and specificity values of the test for each of these comparisons. For the AD versus control comparisons, sensitivity was the proportion of AD patients classified correctly, and specificity was the proportion of controls classified correctly. For the AD versus OD comparison, sensitivity was proportion of AD patients classified correctly, and specificity was the proportion of OD patients classified correctly. For the OD versus controls comparison, sensitivity was the proportion of OD patients classified correctly, and specificity was the proportion of OD patients classified correctly.

A ROC curve is a graphical representation of the relationship between sensitivity, plotted on the *x*-axis, and (1 − specificity), plotted on the *y*-axis. Sensitivity and specificity refer to a test's ability to correctly classify a person as impaired or not impaired. *Sensitivity*, or the true-positive rate (TPR), is the percentage of individuals whom the test classified as impaired who truly are impaired. In Figure 2, this would be defined as a / (a + c). *Specificity*, or the true-negative rate (TNR), is the percentage of individuals whom the test classified as not impaired who truly are not impaired. In Figure 2, this would be defined as d / (b + d).

Thus, (1 – specificity), known as the false-positive rate (FPR), refers to the probability that a test incorrectly classifies a person as impaired who is not impaired.

Each point on the ROC curve represents a different possible cut-off score or decision threshold such that all possible thresholds are shown, and each corresponds to a certain TPR and FPR value (Pepe, Janes, Longton, Leisenring, & Newcomb, 2004). The larger the AUC (i.e., the higher the curve and the larger the AUC value), the more diagnostically accurate the test is. Thus, an AUC value of 1.0 indicates that the test is completely accurate, correctly distinguishing everyone who has the disorder from everyone who does not, whereas an AUC value of 0.0 indicates that the test is completely inaccurate, incorrectly classifying all people with or without the disorder as the opposite (Zhou, Obuchowski, & McClish, 2002). Tests with an AUC value of 1.0 or 0.0 are rare. Practically, AUC values above .50 indicate that the test is better than chance at predicting diagnosis and has at least some discriminative capacity. SPSS calculates the AUC along with its corresponding p value and 95% confidence interval. For a detailed explanation of ROC curves and their utility in diagnostic decision-making, see, for example, Swets, Dawes, and Monahan (2000) or Zhou et al. (2002).

Diagnosis

		Present	Absent	Total
		Present	Absent	Total
Test	Positive	a True Positive	b False Positive	a + b
	Negative	c False Negative	d True Negative	c + d
	Total	a + c	b + d	N = a + b + c + d

Figure 2. Two-by-two table depicting true positive, false positive, false negative, and true negative values from which key diagnostic efficiency statistics are calculated.

Step 6: Diagnostic efficiency statistics. I calculated predictive values using cut scores based on the results of the ROC analyses. I identified cut scores that provided the best balance between sensitivity and specificity such that the highest diagnostic accuracy (i.e., the highest rates of sensitivity and specificity) was achieved. Using these cut scores, I calculated the positive and negative predictive values, overall accuracy or 'hit rate', and likelihood ratios using a 2 (diagnosis) x 2 (test) table similar to that shown in Figure 2.

The *positive predictive value* (PPV) is the percentage of individuals classified by the test as having the condition who really have the condition, which is defined as a / (a + b) in Figure 2 (i.e., true positives / [true positives + false positives]). Conversely, the *negative predictive value* (NPV) is the percentage of individuals classified by the test as not having the condition who truly do not have the condition, which is defined as d / (c + d) in Figure 2 (i.e., false positives / [true positives + false positives]). The *overall classification rate*, also known as the accuracy or 'hit rate', is the proportion of individuals with and without the condition whom the test classifies correctly, and is defined as (a + d) / N in Figure 2.

Whereas the predictive values are dependent on the prevalence of the condition of interest in the chosen population, likelihood ratios are independent of disease prevalence. The *positive likelihood ratio* refers to the likelihood that a positive test result is obtained by a person with the condition of interest, for example, in this case, the likelihood that a person obtaining a certain score has AD. It is calculated as sensitivity / (1 – specificity). The *negative likelihood ratio* refers to the likelihood that a negative test result is obtained by a person with the condition of interest. It is calculated as (1 – sensitivity)/ specificity. I calculated these statistics separately for the AD group and the control group, the AD group and the OD group, and the OD group and the control group.

Step 7. Hierarchical regression: Relationship between BNT-SA-SF performance and dementia severity in AD. To investigate the relationship between BNT-SA-SF performance and AD severity, I conducted a hierarchical regression. I created a regression model, using BNT-SA-SF score as the outcome variable and entering the sociodemographic variables that the GLM (Step 4 above) identified as significant as predictors in the first step, and MMSE score as a predictor in the second step. Dementia research studies use the latter regularly as an alternate to the Clinical Dementia Rating (CDR; Morris, 1993) scale to stage AD. MMSE score has been shown to agree significantly with the CDR ratings of mild, moderate, and severe dementia in patients with AD (Perneczky et al., 2006).

Step 8: Item analysis. To investigate the item-by-item functioning in this sample and to determine the pattern of responses across the groups, I calculated a difficulty index for each item in each of the three groups and for the sample overall. The difficulty index is the proportion of correct responses produced spontaneously or following a semantic cue for an item, and is presented as a percentage. Higher percentages indicate 'easier' items (i.e., items to which many individuals in the sample responded correctly).

Step 9: Normative data. I concluded the analysis by calculating preliminary normative data for the BNT-SA-SF, based on the performance of the control group.

Results

Sample Characteristics

Table 1 presents the sociodemographic characteristics and MMSE performance of the three groups of participants that constituted the final sample. One-way ANOVA detected no statistically significant between-group difference with regard to *age*, $F(2, 117) = 2.80$, $p = .065$, $\eta_p^2 = .046$. Regarding *sex*, chi-square analysis detected no significant between-group difference in the distribution of males and females, $\chi^2(2, N = 120) = 1.64$, $p = .440$, Cramer's $V = .117$. Fisher's exact tests detected no significant between-group differences in the distribution of (a) *race*, $p = .323$, Cramer's $V = .154$, (b) *education*, $p = .070$, Cramer's $V = .214$, or (c) *SES*, $p = .077$, Cramer's $V = .192$. There was, however, a significant between-group difference with regard to *language*, $p = .035$, Cramer's $V = .210$. In order to interpret the significant finding, I inspected the standardized residuals. Although a Fisher's exact test does not utilize standard residuals, the chi-square was also significant for language ($\chi^2(4, N = 120) = 10.33$, $p = .035$), and using standardized residuals to interpret chi-square analyses is a recognized procedure for 3 x 3 designs. None of the standardized residuals were significant (i.e., none were < -1.96 or > 1.96) but it appears that the proportion of Afrikaans and 'bilingual' participants (i.e., individuals who reported they spoke English and Afrikaans equally often) differed between the groups. Specifically, there were more 'bilingual' participants in the control group and more Afrikaans-speakers in the AD group.

To ensure the patient groups were matched with regard to the severity of their cognitive impairment and that the control group was cognitively intact in comparison, a further between-groups comparison examined MMSE scores. The assumptions of normality and homogeneity of variance were violated for these data; unsurprisingly, scores were highly positively skewed for the control group and Levene's test was significant, $F(2, 116) = 29.45$, $p < .001$. In addition, the sample sizes were unequal and therefore, to be conservative, I report Welch's F-ratio (Field, 2009). As expected, there was a significant between-group difference in terms of MMSE scores, Welch's $F(2, 40.459) = 96.738$, $p < .001$, $\omega^2 = .617$. Games-Howell post-hoc procedures confirmed that the control group performed significantly better than the AD group, $p < .001$, and than the OD group, $p < .001$. Although the AD group had slightly lower MMSE scores than the OD group, this difference was not statistically significant, $p = .138$, according the Games-Howell procedure.

Table 1.
Sample Demographic Characteristics

	Group		
	Patient		Control
Variable	AD (*n* = 46)	OD (*n* = 23)	(*n* = 51)
Sex			
Male	13	6	9
Female	33	17	42
Race			
White	10	4	15
Colored	36	17	36
Black	0	1	0
Age (years)[a]			
Range	49.17 - 84.58	50.00 - 84.67	51.00 - 88.25
M (*SD*)	72.25 (8.07)	69.37 (8.82)	67.87 (10.18)
Language[b]			
English	30	17	35
Afrikaans	10	3	3
English/Afrikaans	3	3	13
Education[c]			
≤ 7	13	8	4
8-11	23	10	31
12	6	4	8
College or university degree	4	1	8
SES[d]			
High	5	1	14
Medium	14	8	17
Low	27	13	20
MMSE			
Range	9 - 29	13 - 29	26 - 30
M (SD)	19.78 (4.70)	22.18 (4.77)	28.78 (1.10)

Note. AD = Alzheimer's disease; OD = other dementia; SES = socioeconomic status; MMSE = Mini-Mental State Examination. Due to incomplete patient files, the information regarding language for three participants in the AD group (a 69-year-old male, a 68-year old-female, and a 79-year-old female) was not available; although all indicated they spoke English and Afrikaans, the examiner had not specified which language was spoken most often. The information regarding race for a 72-year-old female participant and the information regarding SES and MMSE for a 69-year-old male, both with VaD and hence included in the OD group, was also not available.
[a]Participant age, in years, was calculated from each participant's age in months at time of testing, which was divided by 12 to achieve a decimal value. [b]Language refers to self-reported language spoken most often. [c]Education is presented in grades successfully completed. [d]High-, medium-, and low-SES brackets were estimated using self-reported monthly income of the participant's household. I classified an income of $\geq$ ZAR5000 ($\geq$ ±US$500) a month as high SES, ZAR2000 to ZAR4999 (±US$200 - ±US$499) a month as medium SES; and < ZAR2000 (< ±US$200) a month as low SES.

Internal Consistency

To estimate the internal consistency reliability of the test, I calculated Cronbach's α, based on scores from the entire sample of participants. Item 1 (*tree*) had a zero variance (i.e., all participants answered this item correctly), and was thus excluded from the reliability analysis. On the remaining 14 items, Cronbach's $\alpha = .80$. No items detracted from the measure's reliability (see Table 2).

Table 2.
Value of Cronbach's α if Item Deleted

Item Number	Cronbach's α if item deleted
1	.80
2	.80
3	.80
4	.80
5	.79
6	.78
7	.77
8	.76
9	.79
10	.77
11	.77
12	.78
13	.79
14	.78
15	.79

Note. Cronbach's α if item deleted displays the value of Cronbach's α if that item were removed from the test.

Diagnostic Capacity of the BNT-SA-SF

Between-Group Differences in BNT-SA-SF Performance. The first step in exploring the diagnostic validity of the BNT-SA-SF was to investigate whether test performance differed between the diagnostic and control groups (see Table 3 for descriptive statistics). Levene's test revealed that the distribution of the data violated the assumption of homogeneity of variance, $F(2, 117) = 12.78$, $p < .001$. Transforming the data did not rectify the problem. Although ANOVA is robust to small violations of this assumption, particularly when samples are equal, the sample sizes in this study were unequal and the data were non-normally distributed (being positively skewed for the control group). Therefore, to be conservative, I report Welch's F. There was a significant between-group difference, Welch's $F(2, 54.80) = 46.033$, $p < .001$, $\omega^2 = .043$. Games-Howell post-hoc procedures revealed that the AD and OD groups performed significantly more poorly than the control group, $p < .001$

in each case, and that the OD group performed significantly better than the AD group, $p <$ .05.

Table 3.
Descriptive Statistics for BNT-SA-SF Performance of the Patient and Control Groups

	AD ($n = 46$)	OD ($n = 23$)	Control ($n = 51$)	Overall ($N = 120$)
Range	2 - 15	7 - 13	10 - 15	2 - 15
M (SD)	9.15 (2.73)	10.61 (1.56)	12.86 (1.23)	11.01 (2.60)

Note. AD = Alzheimer's disease. OD = other dementia.

General linear model: Relationship between BNT-SA-SF performance and sociodemographic variables. Using a general linear model procedure, I performed further analyses to investigate the effect of sociodemographic variables on BNT-SA-SF performance. None of the two-way interactions were significant predictors of performance, and therefore they were removed from the model. After removing all the non-significant demographic predictors (viz., age, education, race, language, and SES) from the model one by one, the only significant predictors of BNT-SA-SF score were sex, $p = .002$ and group, $p <$ .001. Table 4 presents the best-fitting model. The combined influence of sex and diagnostic group explained 48% of the variance in BNT-SA-SF score. For sex, $B = 1.307$, $t = 3.144$, $p =$ .002, indicating that male participants scored on average 1.307 points higher than female participants on the BNT-SA-SF. The BNT-SA-SF was still significant for predicting AD versus control, $B = 3.811$, $t = 9.857$, $p < .001$, and for predicting AD versus OD, $B = 1.473$, $t = 3.024$, $p = .003$.

Table 4.
General Linear Model: Final model predicting BNT-SA-SF (N = 115)[a]

	Type II SS	df	MS	F	p	ESE
Corrected model	355.63	3	118.54	34.41	< .001***	
Sex	34.05	1	34.05	9.89	.002*	.081
Group	340.95	2	170.48	49.49	<.001***	.469

Note. SS = sums of squares; MS = mean square; ESE = effect size estimate (in this case, η_p^2). For the overall model, $R^2 = .48$ (adjusted $R^2 = .47$). [a]Five participants were excluded from the analysis due to missing demographic information (see the note to Table 1 for details of these participants).
*$p < .05$. **$p < .01$. ***$p < .001$.

ROC analyses. To further investigate the discriminative validity of the BNT-SA-SF, I produced ROC curves for (a) the AD group versus the control group (see Figure 3), (b) the

AD group versus the OD group (see Figure 4), and (c) the OD group versus the control group (see Figure 5). Table 5 presents a summary of the ROC analyses and cut scores.

The AUC was significant for all the ROC curves, indicating that the test had a significant discriminatory capacity in terms of sensitivity and specificity for each of the comparisons. The AUC values for the AD versus controls and OD versus controls comparisons were good (between .75 and .92), and the AUC value for the AD versus OD comparison was fair (between .50 and .75). The optimal cut scores for each comparison were selected from the cut-off scores presented in Table 5. I aimed to identify cut scores with the highest sensitivity in detecting cases while attempting to balance specificity, maintaining a low FPR (1 − specificity), for identifying controls accurately; thus, for some comparisons, the balance lies in favor of specificity. The selected cut scores are highlighted in boldface font in the table.
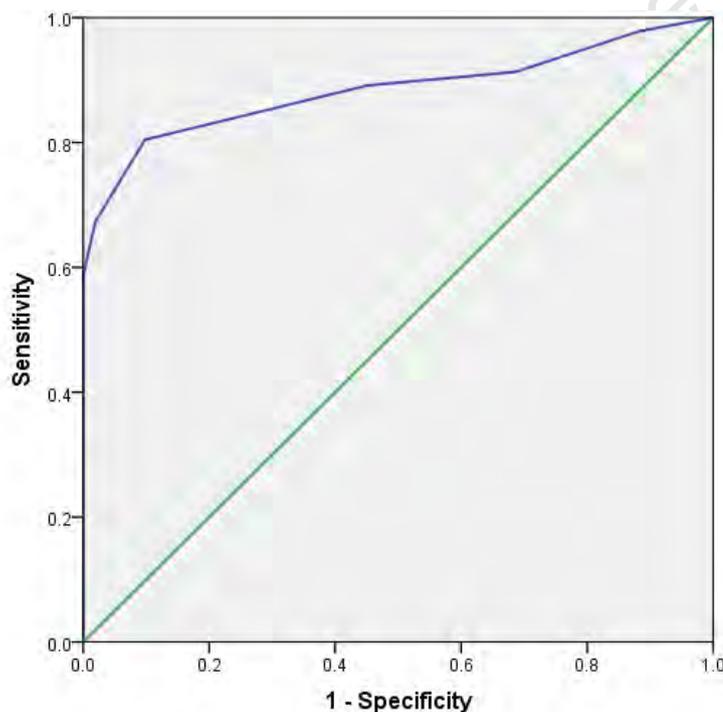
Figure 3. Receiver operating characteristic (ROC) curve of BNT-SA-SF cut-scores for classifying the AD group from controls. The diagonal reference line represents the AUC = .50.
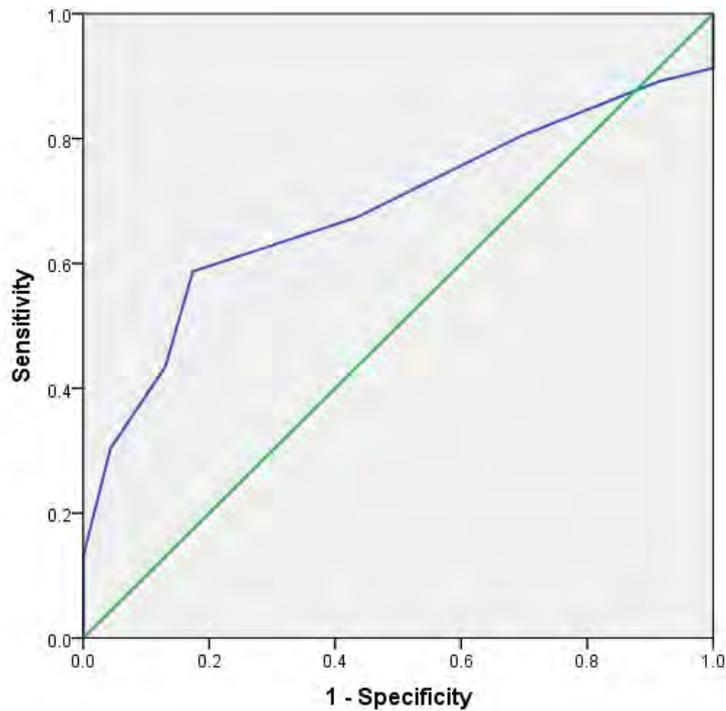
Figure 4. Receiver operating characteristic (ROC) curve of BNT-SA-SF cut-scores for classifying the AD group from the OD group. The diagonal reference line represents the AUC = .50.
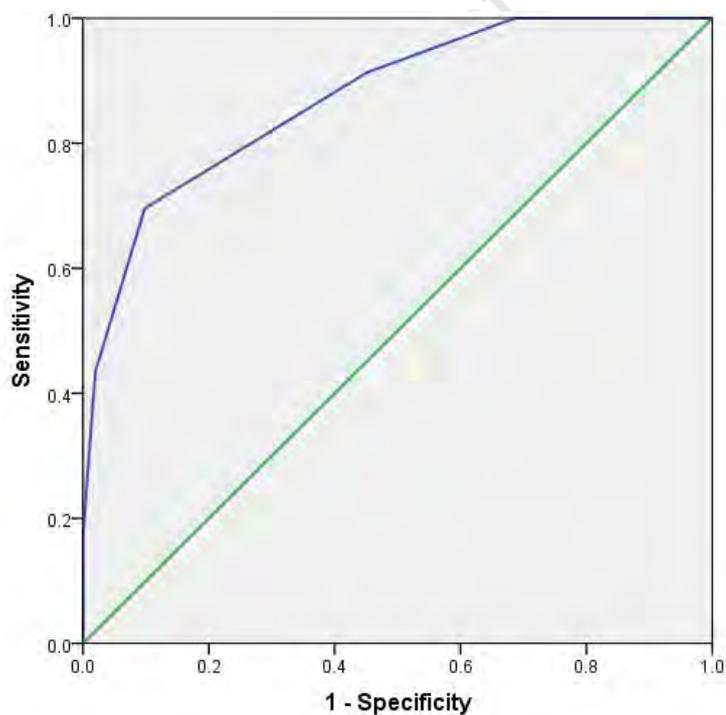
Figure 5. Receiver operating characteristic (ROC) curve of BNT-SA-SF cut-scores for classifying the OD group from controls. The diagonal reference line represents the AUC = .50.

Table 5.
Summary of the Receiver Operating Curve (ROC) Analyses, with Cut-Scores (N = 120)

Comparison	AUC (*SE*)	*p*	95% CI	Cut score	Sensitivity	1-Specificity
AD vs. controls	.884 (.038)	< .001***	.809 - .959	9	.435	.000
				10	.587	.000
				11	.674	.020
				12	**.804**	**.098**
				13	.891	.457
				14	.913	.686
AD vs. OD	.684 (.064)	.013*	.559 - .808	9	.435	.130
				10	**.587**	**.174**
				11	.674	.435
				12	.804	.696
				13	.891	.913
				14	.913	1.000
OD vs. controls	.873 (.044)	< .001***	.787 - .959	9	.130	.000
				10	.174	.000
				11	.435	.020
				12	.696	.098
				13	**.913**	**.451**
				14	1.000	.686

Note. Area under the curve (AUC) is reported with standard error (*SE*) in parentheses. CI = confidence interval. Cut scores identified as representing the best balance between sensitivity and specificity are in boldface font.
*p < .05. **p < .01. ***p < .001.

Diagnostic efficiency statistics. Table 6 shows the optimal cut scores with corresponding sensitivity, specificity, PPV, NPV, overall accuracy or 'hit rate', and the positive and negative likelihood ratios. The values from which these test statistics were calculated for each comparison are presented in Appendix D.

The test discriminated best between the AD group and the control group, with the highest combination of sensitivity and specificity (80% sensitivity, 90% specificity). Using a cut score of 12, the BNT-SA-SF had an overall accuracy of 86% in correctly identifying both positive (patients) and negative (controls) cases. The positive likelihood ratio indicates that, for an individual with a score of 12 or less on the BNT-SA-SF, there is a moderate increase in the likelihood of them having AD. Specifically, a score of ≤ 12 is 8.2 times more likely to occur in a person with AD than a person who is cognitively intact.

The test had a slightly lower, but still reasonable, discriminative capacity between the AD and OD groups (59% sensitivity, 83% specificity). There was a higher degree of overlap between scores for AD and OD patients than in the other comparisons, but the BNT-SA-SF was still able to classify, using a cut score of 10, 67% of patients correctly as having either AD or another type of dementia. A score of ≤ 10 in a patient with dementia indicates a small increase in the likelihood that the dementia type is AD rather than another type of dementia.

Specifically, a score of ≤ 10 is roughly 3 times more likely to occur in a patient with AD than in a patient with another type of dementia.

The BNT-SA-SF also discriminated well between the OD group and the control group (91% sensitivity, 55% specificity). The overall accuracy of the test using a cut-off score of 13 was 56%. A person with dementia, other than AD, is roughly twice more likely to score < 13 than a person who is cognitively intact. The test has a fairly high rate of false positives using a cut score of 13, and thus has a lower specificity value than the other two comparisons. A score of ≤ 13 is roughly twice more likely to be seen in someone who has dementia, other than AD, than someone who is cognitively intact.

Table 6.
Diagnostic Efficiency Statistics for the BNT-SA-SF (N = 120)

	Comparison		
Statistic	AD vs. controls	OD vs. controls	AD vs. OD
Cut-score	12	13	10
Sens. (95% CI)	0.80 (0.66-0.91)	0.91 (0.72-0.99)	0.59 (0.43-0.73)
Spec. (95% CI)	0.90 (0.79-0.97)	0.55 (0.40-0.69)	0.83 (0.61-0.95)
PPV	0.88	0.48	0.87
NPV	0.84	0.93	0.50
Accuracy	0.86	0.58	0.67
+LR	8.20	2.02	3.38
-LR	0.22	0.16	0.50

Note. Sens = sensitivity; CI = confidence interval; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value; +LR = positive likelihood ratio; –LR = negative likelihood ratio. Accuracy refers to the overall classification rate of the test.

Hierarchical regression: Relationship between BNT-SA-SF performance and dementia severity in AD. I performed a further analysis to investigate the contribution of dementia severity to BNT-SA-SF performance in the AD group. I conducted a hierarchical regression analysis, entering sex (the only sociodemographic variable the general linear model in step 4 above identified as influencing BNT-SA-SF performance) and MMSE score

(as a proxy for dementia severity) as predictor variables in separate steps, and test performance as the outcome variable. Table 4 presents the results from this analysis.

Sex significantly predicted BNT-SA-SF score and explained 14% of the variance in the outcome variable. MMSE score was also a significant predictor of BNT-SA-SF score, after partialling out the effect of sex, and explained a further 11% of the variance in the outcome variable. However, the significant effect of sex did not remain significant when MMSE score was added to the model, $p = .072$. Of the two variables, the beta values show dementia severity ($\beta = .36$) has a greater influence on BNT-SA-SF performance than whether the participant was male or female ($\beta = .258$). For every one unit decrease in MMSE score, BNT-SA-SF score decreased by .26. Together, sex and MMSE score accounted for 24% of the variance in BNT-SA-SF score, $F(2, 43) = 7.36$, $p = .002$.

That the effect of sex on BNT-SA-SF performance was not significant when MMSE score was added to the model, suggests sex does not have an independent effect on BNT-SA-SF performance and is partially mediated by dementia severity. This was not substantiated by running a Sobel test, $z = -1.72(0.41)$, $p = .086$.

I checked the assumptions underlying the regression model and the model accuracy. The variance inflation factor (VIF) values were all < 10 and the tolerance statistics were all well above .2, indicating no apparent problems with multicollinearity. In addition, the Durbin-Watson statistic was 2.30, indicating the model residuals were independent. Examination of the plots of standardized residuals and standardized predicted residuals did not detect outliers, heteroscedasity, or deviations from linearity in the data i.e., the assumptions of normality, linearity, and homogeneity of variance were upheld. The Cook's and Mahalanobi's distances were also both within acceptable limits. Overall, then, the model appears reliable and sound.

Table 7
Hierarchical Regression: Predicting BNT-SA-SF score from Sex and MMSE in AD (N = 46)

	B	SE B	β	t	p
Step 1					
Constant	13.02	1.50			
Sex	-2.25	0.84	-.38	8.75	.010*
Step 2					
Constant	7.71	2.51			
Sex	1.55	0.84	-.26	1.85	.072
MMSE	0.21	0.08	.36	2.56	.014*

Note. MMSE = Mini-Mental State Examination. $R^2 = .14$ ($p = .010$) for Step 1; $\Delta R^2 = .11$ for Step 2 ($p = .014$).
*$p < .05$ **$p < .01$ ***$p<.001$.

Item Analysis

To further investigate the characteristics of the test items and the relationship between item difficulty and item placement in the test, I used the performance of the current sample to create a difficulty index for each item of the BNT-SA-SF. I did this for each of the three groups individually (AD, OD, controls) and for all three groups combined (overall).

As Figure 6 shows, the trend for increasing item difficulty as the test progresses was evident. All participants, regardless of group, answered item 1 (*tree*) correctly, whereas very few answered item 15 (*protractor*) correctly. There was a relatively slight but steady increase in item difficulty through to item 12, after which the graph dropped abruptly for all three groups, and overall. On average, participants in all three groups made most of their errors on the last three items (*compass, sphinx,* and *protractor*). Items 13 (*compass*) and 15 (*protractor*) were of considerably greater difficulty than the previous items. Very few people could answer these items correctly. Less than a third of participants in the control group could name either of these items correctly, and none of the 23 participants in the OD group could name item 15 correctly, for example. Item 14 (*sphinx*) was only slightly less difficult, with 37 of 120 participants (31.7%) naming this item correctly. Control group average performance was at or near the ceiling for items 1 through 12, as more than 47 of 51 participants (> 90%) in that group produced a correct response for these items. There was, however, a sharp increase in item difficulty for items 13, 14, and 15. Only 17 (33.3%), 27 (52.9%), and 13 (25.5%) of the 51 control participants produced a correct response for each of these items, respectively.

A number of other observations can be made regarding item difficulty. First, item 2 (*comb*), item 3 (*toothbrush*), and item 4 (*hanger*) appear to be of approximately equal difficulty; the same proportion of correct responses were produced for each item. Second, item 12 (*funnel*) appears to be slightly easier than the previous two items. Overall, 98 of 120 participants (81.7%) answered *funnel* correctly, compared to 87 (72.5%) and 71 (59.2%) who answered the previous two items correctly. Item 9 (*dominoes*) also appeared to be slightly easier than the previous two items for the AD and OD groups, but not for the control group.

Also of relevance here is that participants in the AD group made more errors across all the items than did participants in the OD group and control groups; naturally, this pattern of performance was detected by the between-group comparisons reported above. However, participants in the OD group made more errors than those in the AD group on items 14 (*sphinx*) and 15 (*protractor*).
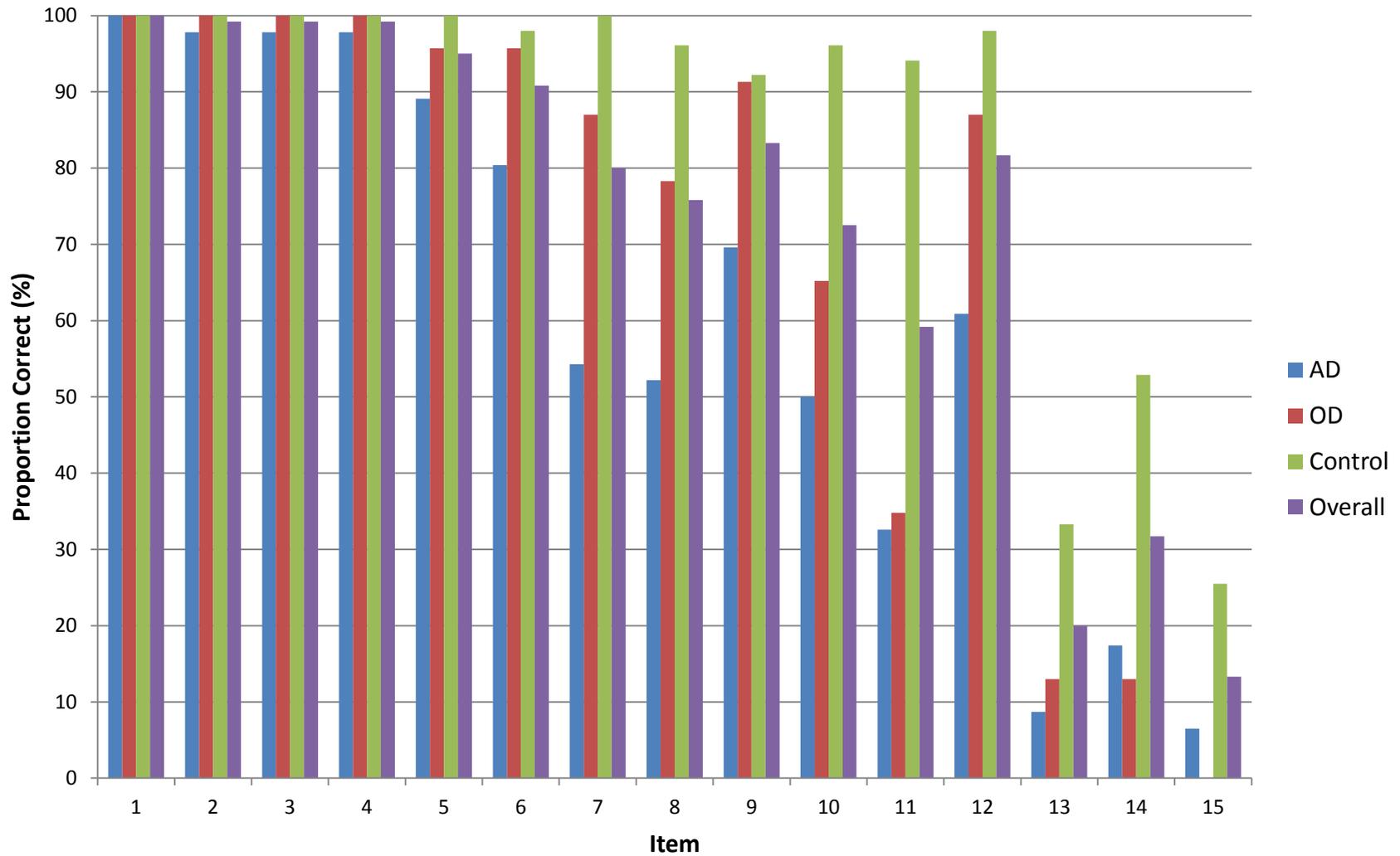
Figure 6. Difficulty index displaying the proportion of correct responses made spontaneously or with semantic cue for AD, OD, and control participants, and overall for the entire sample.

Preliminary Normative Data

Tables 8 and 9 present preliminary normative data for the BNT-SA-SF, based on the performance of the present study's cognitively intact participants. For reasons grounded in (a) the analyses described above, (b) inspection of the data, and (c) a review of the literature, the normative data are stratified according to sex and level of education (0 - 7 years, 8 – 12 years, > 12 years). Table 8 presents the norms for men and women. Table 9 presents the norms stratified by sex and education.

Table 8.
Normative Data for the BNT-SA-SF Stratified by Sex

	Sex	
	Male	Female
n	9	42
M (*SD*)	13.89 (1.05)	12.64 (1.17)

Table 9.
Normative Data for the BNT-SA-SF Stratified by Sex and Level of Education

	Sex			
	Male		Female	
Education (years)	*n*	*M* (*SD*)	*n*	*M* (*SD*)
0 - 7	-	-	4	11.50 (1.00)
8 - 12	7	14.00 (0.82)	32	12.69 (1.06)
> 12	2	13.50 (2.12)	6	13.17 (1.47)

Note. Education is presented in grades completed. Those who fall into the > 12 category completed either a college or university degree.

Discussion

Distinguishing Alzheimer's disease (AD) from normal aging and other types of dementia is an important, but challenging, clinical undertaking. The Boston Naming Test (BNT) is a popular confrontation naming task, and studies have shown short and full versions of the test to be particularly useful in detecting the naming deficits present in AD. Despite the diagnostic utility of the BNT, it is difficult for clinicians to interpret the performance of patients whose demographic backgrounds differ from those individuals who constitute the standard normative sample (Kohnert et al., 1998). The test was developed for the assessment of English-speaking, North American individuals; it is not surprising, then, that BNT performance reflects a cultural bias when used for the assessment of individuals who are not first-language English speakers and/or who are not North American. As Kohnert et al. (1998, p. 424) states, "One cannot simply assume that the BNT is a valid measure for any populations other than the one for which that instrument was first developed and normed."

The present study assessed the diagnostic utility of a South African short form adaptation of the BNT, using a well-defined population of patients with dementia from a Cape Town memory clinic and healthy, community-dwelling control participants. Specifically, I attempted to validate the use of the BNT-SA-SF for distinguishing AD patients from controls and from patients with other types of dementia. To clarify the diagnostic validity of the set, I set out to address five main questions:

(1) What is the internal consistency reliability of the BNT-SA-SF?

(2) How well can it distinguish between (a) patients with AD and controls, (b) patients with AD and patients with dementia diagnoses other than AD, and (c) patients with dementia diagnoses other than AD and cognitively intact controls?

(3) Are there real between-group differences in BNT-SA-SF performance, or do sociodemographic variables influence performance on the test?

(4) Is there a relationship between BNT-SA-SF performance and dementia severity (estimated by the Mini-Mental State Examination; MMSE) in AD?

(5) What is the pattern of performance, in both patients and controls, across the 15 test items?

I now discuss how the observed data and subsequent statistical analyses answered each of those questions.

Implication of Results

The sample consisted of 120 older adults divided into three diagnostic groups (AD, other dementia (OD), and controls). Individuals of varying ages, education levels, SES brackets, races, and languages comprised the sample, and there were considerably more women than men. Although the three groups were well matched on most sociodemographic variables, they differed significantly in terms of the distribution of Afrikaans-speaking participants (who reported they spoke Afrikaans most often) and 'bilingual' participants (who reported they spoke English and Afrikaans equally often).

Reliability. The internal consistency of the BNT-SA-SF (Cronbach's alpha = .80) compares favorably with that reported in the literature for the full-length BNT and for various short forms. Reported values of Cronbach's alpha range from .78 to .96 for the 60-item test, from .57 to .90 for 30-item short forms, and from .31 to .84 for 15-item short forms (Fastenau, et al., 1998; Graves et al., 2004; Saxton, Ratcliff et al., 2000; Tombaugh & Hubley, 1997; B. Williams et al., 1989). Hence, this is a positive result for the BNT-SA-SF, considering that the internal consistency of a test is usually affected by the number of items in the test, with longer tests typically obtaining higher values than shorter ones.

Validating the diagnostic capacity of the BNT-SA-SF. A number of analyses assessed the diagnostic capacity of the test. In interpreting the results of these analyses, it must be noted that the BNT-SA-SF, in its full or shortened form, is a *screening* measure, not intended for use as a diagnostic tool to make final clinical diagnoses. A screening test is traditionally used as part of a larger battery of neuropsychological tests, the results of which clinicians or researchers use to identify the presence or absence of impairment across a range of cognitive domains. The BNT thus serves to identify impairment in visual confrontation naming. This identification then acts as a basis or point of departure for further investigation, rather than establishing the presence or absence of disease (e.g., AD). Beyond a patient's performance on a neuropsychological test battery, clinicians also refer to the patient's medical history, biographical information, and the results of other relevant medical tests, such as MRIs (Kalula et al., 2010; McKhann et al., 1984).

There was a significant between-group difference in BNT-SA-SF performance. A series of pairwise post-hoc comparisons confirmed the a priori hypotheses that (a) patients with AD will perform significantly more poorly than cognitively intact controls, (b) patients with AD will perform significantly more poorly than patients with dementia diagnoses other than AD, and (c) patients with dementia diagnoses other than AD will perform significantly more poorly than cognitively intact controls. The confirmation of the first hypothesis, (a), is

consistent with previous studies reporting that various BNT short forms are able to distinguish patients with AD from healthy controls based on differences in total score (Calero et al., 2002; Jefferson et al., 2007; Lansing et al., 1999; Mack et al., 1992; B. Williams et al., 1989).

The confirmation of the second, (b) and third hypotheses, (c), suggests that the BNT is able to distinguish patients with other types of dementia from normal aging and to distinguish patients with AD from patients with other dementia diagnoses based on total score. With regard to the literature, studies have shown that the BNT and other confrontation tasks are able to differentiate patients with dementia from controls, where patients with different types of dementia form a single dementia group, (Brouillette et al., 2011; de la Plata et al., 2009; Miller et al., 2010). Both of these findings are consistent with the finding reported in the literature that patients with vascular dementia (VaD), who comprised 65.2% of the OD group in the present study, exhibit confrontation naming impairment, but make less naming errors than patients with AD (Barr et al., 1992; De Jager et al., 2003; Schmidtke & Hüll, 2002). There are limited and inconsistent findings regarding differences in BNT performance between patients with AD and those with alcohol-related dementia, Lewy Body dementia, Parkinson's disease with dementia, or frontotemporal dementia, the other dementia diagnoses that make up the OD group; however, some studies have shown that patients with some of these dementia types perform more poorly than patients with AD (Bayles & Tomoeda, 1983; Diehl et al., 2005; Grossman et al., 2004; Noe et al., 2004; Saxton, Munro et al., 2000).

In addition, it is important to interpret the difference in BNT-SA-SF performance between the AD and OD groups while taking into consideration that the AD group scored slightly below the OD group on the MMSE ($M = 19.78$ vs. $M = 22.18$). Even though this difference was not statistically different, it suggests that the AD group may have been more impaired than the OD group overall, and that this may account for their poorer performance on the BNT-SA-SF rather than the a more substantial naming impairment in AD relative to other dementia types. However, the MMSE is sensitive to memory and language impairment, which are some of the primary deficits seen in AD, and therefore AD patients may be rated as disproportionately more impaired than patients with other types of dementia, which have different primary deficits (Kramer et al., 2003).

Relationship between BNT-SA-SF performance and sociodemographic variables. The AD, OD, and control participants were well matched on most key demographic variables that have been shown to affect performance on the BNT. However, it was important to ensure

that the underlying naming deficit, rather than any differences in sociodemographic characteristics, accounted for the observed between-group differences in BNT-SA-SF performance. It is therefore particularly notable that the general linear modeling (GLM) analysis revealed that group status was the most significant predictor of BNT-SA-SF performance, over and above any of the sociodemographic variables. Although there was a significant relationship between sex and BNT-SA-SF score, with male participants scoring significantly higher than female participants, group status was still a significant predictor of test performance. In other words, BNT-SA-SF performance could still distinguish AD from normal aging, even after the effects of sex were accounted for. The GLM results confirmed that variations in age, education, language, race, and SES do not undermine the ability of the BNT-SA-SF to differentiate AD from normal aging or AD from other types of dementia. This pattern of results suggests that between-group differences in test score can be attributed to varying severity or presence of an underlying naming impairment.

The only significant sociodemographic predictor of BNT-SA-SF performance was sex. This finding is somewhat inconsistent with the literature, as most BNT studies have not found a significant effect of sex on confrontation naming performance (Cruice et al., 2000; Fastenau, 1998; Ross et al., 1995; Saxton, Ratcliff et al., 2000). However, some studies have found that in older samples particularly, men tend to perform better on the BNT than women (Randolph et al., 1999; L. Welch et al., 1996). Those studies that have found an effect of sex have generally attributed this to an interaction between sex and education in that men in older cohorts tend to have higher levels of education than women (Jefferson et al., 2007; L. Welch et al., 1996). In the present study, there was no significant interaction between these age and education, however. The significant effect of sex on BNT-SA-SF may be explained by a recent study that found men outperform women on the BNT in healthy older adult and AD samples, after taking the effects of age, IQ, and education into consideration, and thus the sex differences found in older samples represent a genuine effect (Hall et al., 2012).

It is important to note that there were considerably fewer men (*n* = 28) than women (*n* = 92) in the current sample. It is unclear why there were more women than men in the final sample of patients with dementia from the Groote Schuur Hospital Memory Clinic. As I recruited control participants to match the Memory Clinic sample with regard to sociodemographic variables, including sex, there were a small number of males in both the patient and control samples. The proportion of men in the patient sample is not representative of the overall profile of patients referred to the Memory Clinic; Kalula et al. (2010) report that 61% of the patients presenting at the Memory Clinic between 2002 and 2008 were

women. It is also not representative of the proportion of men and women in South Africa; 16% of South Africa's population is over 50 years of age and of these, just over half (57.5%) are women (Statistics South Africa, 2012). However, the literature reports an increased risk for dementia in women, compared with men, particularly with regard to AD (Andersen et al., 1999; Gao, Hendrie, Hall, & Hui, 1998). Although some studies do not report such differences (Katz et al., 2012; Ruitenberg, Ott, van Swieten, Hofman, & Breteler, 2001). Nonetheless, clinicians or researchers using the BNT-SA-SF must take the sex of the patient into consideration when assessing their performance.

It is particularly significant that age, education, language, race, and SES were not significant predictors of BNT-SA-SF performance. These findings contradict the conclusions drawn in the extant literature, and by practicing clinicians in South Africa, that it is vital to consider demographic variables such as SES and education when assessing patient performance on neuropsychological tests (Manly, Byrd, Touradji, & Stern, 2004; Nell, 2000; Shuttleworth-Edwards, Kemp et al., 2004; Strauss et al., 2006). Despite the fairly small, but consistent, age effect reported in the literature, with performance decreasing after the age of 70 years (Au et al., 1995; Mitrushina & Satz, 1989; Zec et al., 2005), the present study found no significant effect of age on BNT-SA-SF performance. Once again, this is suprising considering that old older adults were well represented; 64 were > 70 years of age and 17 of these were > 80 years of age.

Similarly, numerous studies have shown an effect of education, particularly at lower levels, on BNT score (Hawkins & Bender, 2002; Le Dorze & Durocher, 1992; Neils et al., 1995). Those studies that have failed to find significant education effects have typically included only individuals with 12 or more years of education (Fastenau, 1998). Despite the current sample having a larger proportion of individuals with lower levels of education (≤ 7 years and 8-11 years) than many other BNT studies, level of education did not compromise the discriminative ability of the test. Although surprising, this result is consistent with previous studies, which have shown that the BNT can distinguish dementia even in low-education samples. For instance, Calero et al. (2002) found that the full 60-item test and a new 15-item version developed by the authors were able to distinguish patients with dementia from healthy elderly in a sample of elders in Spain, even though very few individuals in that sample had education beyond the primary school level and nearly half were classified as functionally illiterate.

Beyond level of schooling completed, quality of education also has the potential to influence test performance (Foxcroft, 2004; Manly et al., 2004; Manly, Jacobs, Touradji,

Small, & Stern, 2002; Shuttleworth-Edwards, Kemp et al., 2004). In South Africa, a history of legislated segregation means that, in older adult samples, some racial/ethnic groups will have received a lower quality of education than others. Specifically, the Apartheid system (1948–1994) created disparities in the quality of education received by South Africa's different race groups. The Bantu Education Act of 1952 ensured that schools serving Black and Coloured (mixed ancestry) learners received a different curriculum, aimed at keeping these population groups within the working class, and less funding than schools serving White learners. Hence, there were higher teacher-to-student ratios and fewer resources in Black and Coloured schools than in White schools. Because participants in the present study were educated during the Apartheid era, race-based differences in the quality of education received by Coloured and White participants might have thus been expected to differ. Specifically, the former may have received a poorer quality of education than the latter. However, race was not a significant predictor of BNT-SA-SF performance, suggesting that (a) BNT-SA-SF performance is not unduly influenced by an individual's quality of education received, or (b) race is a poor proxy for quality of education and a more accurate measure is needed. Alternative measures include reading level (Manly et al., 2004) or type of school attended (Shuttleworth-Edwards, Kemp et al., 2004).

Although there are no published studies that have found a significant effect of SES on BNT performance[4], researchers are essentially obliged to include SES as a variable of interest in studies conducted in lower middle income countries (LMICs) such as South Africa. This obligation stems, largely, from the high rate of economic inequality in such countries. The World Bank (2012) lists South Africa's Gini coefficient, a measure of inequality, as 63.1, one of the highest in the world. Studies that have examined the effects of SES on other cognitive tests report that low SES has a direct effect on performance of test of various cognitive domains (Noble, McCandliss, & Farah, 2007; Noble, Norman, & Farah, 2005) and affects the relative influence of other variables, such as education, on performance on cognitive tests (Dotson, Kitner-Triolo, Evans, & Zonderman, 2009). Although SES is notoriously difficult to quantify, particularly in LMICs (Myer, Stein, Grimsrud, Seedat, & Williams, 2008), in the present study I used participants' average monthly income as a measure of SES. This was the best indicator of SES from the data that was available for Memory Clinic patients, but it is possible that a more accurate measure of SES, including more economic indicators, may have

---

[4]See La Barge et al. (1992) for a study that did not find a significant relationship between SES and BNT performance.

allowed the statistical analyses to detect a significant effect where the current measure did not.

The result that language was not a significant predictor of BNT-SA-SF performance stands in contrast to previous studies suggesting that language experience or proficiency may influence older adult performance on the BNT and other cognitive tests (Bialystok & Craik, 2007; Bialystok, 2007; Gollan et al., 2007; Roberts et al., 2002). However, the results are not necessarily surprising, for two reasons. First, all participants reported they spoke English, even if this was not their first or home language. Second, the BNT-SA-SF is translated into the three main languages spoken in the Western Cape, namely English, Afrikaans, and isiXhosa, and the scoring sheet (see Appendix B) is presented in a multilingual format. Hence, correct answers given in any of these three languages are accepted, irrespective of the fact that the neuropsychological tests at the Groote Schuur Hospital Memory Clinic are usually administered in English, by English-speaking neuropsychologists, to participants who are fluent in English (even if this is not their home language).

That answers in English and Afrikaans were scored as correct in the present study, speaks to the dual-language scoring benefit identified in the literature with English-Spanish bilinguals. When scores are composed of the total number of correct responses irrespective of language, balanced bilinguals, who are equally proficient in both languages, perform significantly better than when scored for a single language (Gollan et al., 2007; Kohnert et al., 1998). Furthermore, Kohnert et al. (1998) found that first or home language did not necessarily remain as the strongest language with regards to BNT performance. In this study, an English-Spanish bilingual sample, for whom Spanish was their home language, scored significantly better on the BNT when tested in English than in Spanish.

In South Africa, where English is the de facto medium of instruction in schools and language of communication in the workplace, individuals may feel more comfortable responding to surveys or tests in this language, rather than in their home language. A recent study examining the career aspirations of 274 Black isiXhosa-speaking school learners found that the majority preferred to complete the measure in English or using a combination of English and Xhosa; only a small number ($n = 42$) responded solely in isiXhosa, even though this was their first or home language (Watson, McMahon, Foxcroft, & Els, 2010). Those findings echo Foxcroft (2004) in recommending that test developers in South Africa produce tests that include various language versions in a either a bilingual or multilingual format. Thus, the multilingual format of the BNT-SA-SF score sheet appears beneficial and can help to reduce the likelihood of language effects on BNT-SA-SF performance.

In summary, these nonsignificant findings for most of the sociodemographic variables suggests the BNT-SA-SF's ability to differentiate dementia is relatively immune to the detrimental effects of these variables in the current sample (Shuttleworth-Edwards, Donnelly, Reid, & Radloff, 2004). The results of this aspect of the data analyses provide promising evidence that the BNT-SA-SF is able to differentiate AD from normal aging in South African samples even if individuals are of different race, have a limited level of education, speak a language other than English as their first language, or are from lower SES backgrounds. In addition, if race is taken as a proxy for quality of education, the BNT-SA-SF is able to identify the poor naming impairment in AD in individuals who have received education of varying quality. However, clinicians must still be careful to consider the demographic profile of the patient, particularly as South Africa's population is so diverse many patients may not fit the demographic profile of the participants in the present study.

Sensitivity, specificity, and predictive values of the BNT-SA-SF. The between-groups comparison showed clear differences in BNT-SA-SF performance. However, such analyses can overestimate the ability of a test to discriminate between diagnostic categories (De Jager et al., 2003). The ROC analyses were more revealing regarding the ability of the test to discriminate between the groups as they allowed (a) investigation of the balance between optimal sensitivity and specificity, and (b) generation of diagnostic validity indices.

The results of ROC analyses confirmed the hypothesis that the test had the strongest diagnostic utility in the differentiation of patients with AD from healthy controls, as compared to the differentiation of (a) patients with dementia diagnoses other than AD from healthy controls, and (b) patients with AD from patients with other dementia diagnoses. The AUC value, a measure of the ability of the test to differentiate between the groups under investigation, was the largest for the AD versus controls comparison (.884). It was slightly smaller, though still good, for the OD versus controls comparison (.873). The AD versus OD comparison had the smallest, though still significant, AUC value (.684).

The ROC analyses identified cut-scores that differentiated AD patients from controls and from patients with other types of dementia, as well as patients with other types of dementia from controls. Standard scoring dictates that cut-scores are 2 *SD*s below the mean, and studies often use this benchmark (Jefferson et al., 2007; Kavé, 2005). Heaton et al. (1999) suggest that a 1 *SD* cutoff for the BNT provides the best balance between sensitivity and specificity. Using ROC analyses to identify cut-scores is more informative, however, as they show the relationship between sensitivity and specificity for all possible cut-scores.

Because the BNT-SA-SF is typically used as a screening measure, optimal cut-scores were chosen to maximize sensitivity, so as to identify all those who may possibly have dementia. The sensitivity of the test was weighed against the specificity in making this decision. For instance, using a cut-score of 13 would have increased the sensitivity of the test in classifying AD from controls by 9% (from 80% to 89%); however, this was weighed against the significant reduction in specificity of 36% (from 90% to 54%) that would have resulted in a large number of false positives for cognitively intact controls. Thus, I decided that 12 would be the best cut-score for differentiating AD from controls. Similarly, with regard to the AD versus OD comparison, after weighing up the sensitivity and specificity values associated with the various cut-scores, the selected cut-score had a higher specificity (83%) than sensitivity (59%). Although high sensitivity is important in a screening measure such as the BNT-SA-SF, the lower sensitivity here is acceptable as a high specificity between AD and other types of dementia means that most dementia patients without AD do achieve a negative test result (> 10). Conversely, a person with a positive test result (< 10) is more likely to have AD. The lower sensitivity means false negatives are also more likely, however.

In summary, when differentiating AD patients from normal controls using a cut-score of 12, the sensitivity and specificity values were good. When differentiating AD from other types of dementia in the OD group using a cut-score of 10, the sensitivity value was somewhat lower, but the specificity of the test was still good. When differentiating other types of dementia from normal controls using a cut-score of 13, the sensitivity of the test was good and the sensitivity was lower, but still fair.

Although the sensitivity and specificity values provide useful information, the positive and negative predictive values provide information that is more relevant for clinicians who may use the BNT-SA-SF with individuals suspected of having dementia. The PPV and NPV of a test allow the clinician to estimate the probability of the condition in an individual patient based on their test result (Kessel & Zimmerman, 1993). With regard to discriminating between AD patients and controls, the PPV indicates that a person who scores 12 or less on the BNT-SA-SF has an 88% probability of truly having AD. The NPV indicates that a person who scores > 12 has an 84% probability of truly not having AD (or a 16% probability of having AD). With regards to discriminating between AD and other types of dementia, the PPV indicates that a person with dementia who scores 10 or less on the BNT-SA-SF has an 87% probability of truly having AD. The NPV indicates that a person with dementia who scores > 10 has an equal probability (50%) of having either AD or another type of dementia. With regards to discriminating between other types of dementia and

controls, the PPV indicates that a person who scores 13 or less on the BNT-SA-SF has a 48% probability of truly having a type of dementia other than AD. The NPV indicates that a person who scores > 13 has a 93% probability of truly not having a type of dementia other than AD (or a 7% probability of having another type of dementia).

These diagnostic efficiency statistics confirm the hypotheses further, and show that the BNT-SA-SF is able to discriminate well between AD and normal aging, but is not as efficient in discriminating between AD and other types of dementia, and other types of dementia and controls.

Relationship with dementia severity in AD. The results of the regression analysis showed that MMSE score is a significant predictor of BNT-SA-SF performance, over and above the variance accounted for by sex. The significant relationship between MMSE and BNT-SA-SF performance suggests the BNT-SA-SF is sensitive to dementia severity in AD and BNT-SA-SF scores are inclined to decrease as dementia severity increases.

A number of studies have found that BNT performance declines with the progression of AD as patients become more severely impaired (Faber-Langendoen et al., 1988; LaBarge et al., 1992; Price et al., 1993). The Clinical Dementia Rating (CDR; Morris, 1993) scale is often used as the 'gold standard' for identifying where a patient lies along the spectrum of mild, moderate, and severe dementia. Although reliable and valid, it is time-consuming to collect sufficient data from the patient and collateral to calculate a CDR score. Thus, the MMSE has sometimes been used as an alternate to the CDR for the staging of AD in dementia research (Chosak Reiter, 2000; Kramer et al., 2003; Larrain & Cimino, 1998). Although the MMSE is, essentially, a measure of global cognitive impairment, scores on the instrument have been shown to correlate with the CDR ratings of mild, moderate and severe dementia in patients with AD (Perneczky et al., 2006). Hence, it is useful for South African clinicians to be aware that BNT-SA-SF performance declines along with MMSE score in patients with AD.

Item analysis. The value of the BNT-SA-SF lies in the fact that its items are drawn from the pool of items that comprise the original test, whereas many other modified versions of the BNT have replaced items from the original test with alternate items. That the items are drawn from the original test has two main advantages. First, it means that an entirely new test does not need to be developed, and that individuals already in possession of the standard BNT will be able to use the modified short form rather than having to purchase new materials. Otherwise stated, the BNT-SA-SF is a cost-effective option, which is a particularly important consideration when one operates in resource-limited settings such as South Africa.

Second, the fact that the items of the BNT-SA-SF are drawn from the original test is conducive to cross-linguistic or cross-cultural comparisons. Extensive adaptations, where many of the original BNT pictures are removed and replaced, particularly when the replacement items are culture- or language-specific, prevents score comparisons across studies.

To address the fourth main question set out in the study aims, regarding the pattern of performance across the test items, I calculated a difficulty index for each item of the test. The difficulty index revealed the overall trend across the 15 items was for increasing difficulty with the progression of the test. However, a number of items appeared out of sequence. Some were easier than their position in the test suggested, whereas others were more difficult than their position suggested. For example, item 12 (*funnel*) appeared to be relatively easier for South Africans than adjacent items, particularly item 11 (*stethoscope*).

Items 2 (*comb*), 3 (*toothbrush*), and 4 (*hanger*) were relatively easy items, and most participants in both the patient and control groups were able to name these items correctly. Moreover, these items performed similarly in each group; each produced the same proportion of correct responses for the three items. If the test needed to be shortened, these items would be candidates for removal as they are of equal difficulty, and therefore provide redundant psychometric information. One argument for retaining such items is that, although it is important to have items of varying difficulty in any version of the BNT, easier items are important for building a patient's confidence, allowing patients to become comfortable with how the test functions, and for building rapport between the patient and examiner.

Conversely, participants in all three groups found items 13 (*compass*), 14 (*sphinx*), and 15 (*protractor*) particularly difficult. Although the BNT is intended to feature items of increasing difficulty, with less familiar items towards the end of the test, very few participants, even cognitively intact controls, were able to name these items correctly, either spontaneously or following a semantic cue. Control participants sometimes indicated that they knew what these items were but did not know the names. For example, a number of participants indicated they knew the items *compass* and *protractor* were instruments used in school or for mathematics and that the item *sphinx* was from Egypt, but they did not know what the correct names were. A plausible explanation for the poor performance on the items *compass* and *protractor* is that participants only encountered or used these items during their formal schooling, which would have taken place at least 40 or 50 years ago. For those with a poor quality or level of education, they may have never used a compass or a protractor. Similarly, regarding the item *sphinx*, whereas well-educated South Africans may be expected

to know this item even though it is not commonplace in South African culture or environment, others, particularly those with a lower level and quality of education, may not be familiar with this item at all.

The item *protractor* is frequently identified as one of the most difficult items in other BNT studies. Studies from various countries have reported a similar difficulty index to the one found in the present study; those studies have included samples of older adults from Australia (27.9% named the item correctly; Worrall et al., 1995) and young adults from New Zealand (27.6%; Barker-Collo, 2001). Even in North America, the item has a low rate of correct responses among cognitively intact individuals. In a Canadian sample of 25-88-year-olds, only 39.7% were able to name this item correctly either spontaneously or following a semantic cue. In the US, Barresi et al. (2000) found that *protractor* was one of a handful of items that were particularly problematic in their sample of community-dwelling adults aged 50-79 years. In the latter study, participants frequently did not name this item correctly spontaneously or following a semantic cue at three test sessions across a 7-year period. The authors suggest that this is an example of a word that was either not well learnt or that was never part of the person's vocabulary. This same reasoning may apply to the BNT-SA-SF items on which control participants performed particularly poorly. If even healthy control participants do not know these items, then it is difficult to draw conclusions that poor performance by a cognitively impaired individual is an indication of a genuine naming impairment; rather, it might be that these items were minimally learnt or were never part of the individual's vocabulary to begin with.

Despite the literature showing that *protractor*, *compass*, and *sphinx* are fairly difficult items regardless of culture and language, the current sample still performed more poorly on *compass* and *sphinx* particularly, than samples from other English-speaking countries. For example, all the participants in a sample of 20 healthy older adults in New Zealand, albeit with a high average level of education (13.9 years), were able to name *compass* and *sphinx* correctly (Barker-Collo, 2007). Correct response rates of over 70% are consistently reported for the item *sphinx*, and of over 60% for the item *compass*, in other English-speaking countries (Barker-Collo, 2007; Tombaugh & Hubley, 1997; Worrall et al., 1995). In contrast, only 52.9% (27 out of 51) and 33.3% (17 out of 51) of the controls in the present study knew the correct names for each of these items respectively. In another South African study, Mosdell et al. (2010) found that the item *sphinx* was one of the most problematic items of the B. Williams et al. (1989) 30-item odd-numbered test, with 22 incorrect answers in a sample of 30 cognitively intact English-, Afrikaans-, and isiXhosa-speaking individuals. These

results suggest these items may be culturally biased or not appropriate for the South African context.

On the other hand, some BNT-SA-SF items may be more familiar to South Africans than to individuals from other countries. With regard to the item *rhinoceros* (item 8 of the BNT-SA-SF), the difficulty index for the entire sample was 75.8%. This number was, however, undoubtedly affected by naming impairment present in the patients with dementia, given that the difficulty index for the control participants was very high. In the control group, 49 of the 51 participants (96.1%) were able to name *rhinoceros* correctly. In comparison, in a Canadian sample of 219 well-educated (*M* = 12.9 years; *SD* = 2.3) 25-88 year olds, the difficulty index for the item *rhinoceros* (item 31 in the 60-item BNT) was 90.4% and it was more difficult than its placement in the test suggested (Tombaugh & Hubley, 1997). In an Australian sample of 136 cognitively intact 57-92 year olds, 88.2% of participants were able to name this item correctly (Worrall et al., 1995).

Of pertinence here is that the rhinoceros is an indigenous animal in South Africa, and is part of the country's cultural heritage. The rhinoceros, or 'rhino' as it is commonly referred to, has received increased and widespread attention in the South African media since 2011 due to awareness-raising around rhino conservation efforts in response to the prolific poaching and threat of extinction of these animals in South Africa and neighboring countries. Thus, South African participants may be more likely to name this item correctly (a) than individuals from other countries, and (b) since 2011 than in the past, due to having encountered it on television, radio, or in other news media.

Preliminary normative data. The focus of this study was not to provide normative data. However, as no published normative data are available for any BNT versions modified for use in South Africa, and because normative data based on samples tested in other countries might not be appropriate, I provided preliminary normative reference data based on the performance of the control group, and stratified by sex and education. Because the control group was matched with the patient samples on key defining sociodemographic variables, these normative data are applicable to patients who may visit Groote Schuur Hospital (and, specifically, those referred to the Memory Clinic). Therefore, clinicians at the Hospital (and in the Western Cape province, and possibly in South Africa) should be able to apply this normative dataset to patients of similar sociodemographic profiles, thus aiding in the more accurate assessment of those individuals. It is important to note, however, that the sample size is small, and that these norms are preliminary and should therefore be used with caution.

Limitations and Directions for Future Research

There were several limitations to this study that need to be discussed in interpreting the study results. First, one should consider the possibility that the spectrum of clinical severity of the dementia patients included in the patient groups may have affected the sensitivity and specificity of the test. Whereas some studies have shown confrontation naming is significantly impaired early in the disease process in AD (Chen et al., 2001), others have found it to be significantly impaired only as patients become more severely demented (Bayles & Tomoeda, 1983). Thus, the test might have greater sensitivity for detecting patients given a diagnosis of 'probable AD', who are in later stages of the disease, than patients with 'possible AD', who are in earlier stages of the disease, or than patients where the etiology is unclear (e.g., in mixed AD/VaD). Although the current study demonstrated the BNT-SA-SF is sensitive to level of cognitive impairment reflected in MMSE performance, it did not directly compare the performance of mild, moderate, or severely impaired groups. Future studies should determine the ability of the test to distinguish mild, moderate, and severe AD from normal aging and from mild, moderate, and severe cases of other dementia types. With much research interest currently focused on differentiating early AD from mild cognitive impairment (MCI) and from normal aging (Albert et al., 2011; Balthazar et al., 2008; Joubert et al., 2010; Morris et al., 2001) such comparisons would be of particular interest.

Second, patients with a variety of dementia subtypes comprised the OD group. Although this group performed more poorly than healthy controls and slightly better than AD patients, as hypothesized, the fact that the group was heterogeneous is limiting. It does not allow decisive conclusions to be drawn regarding the discriminative capacity of the test with regards to 'dementia types other than AD', as this category is broad and not all dementia types were represented in the OD group.

Third, the test results could be skewed by the inclusion of patients with vascular dementia in the two different clinical groups (recall that patients with mixed AD/VaD were included in the AD group, and patients with pure vascular dementia were included in the OD group). Hence, in light of the second and third points, mentioned above, future research should seek to include larger samples and separate groups for each dementia type. For example, future studies could compare the performance of a pure AD group, a mixed AD/VaD group, and a pure VaD group, allowing more focused analysis. Alternatively, separate samples could also be included for the other dementia subtypes, so that a study might include a Lewy Body dementia group, an HIV-associated dementia group, and so forth.

Fourth, it is important to note, that although the experienced medical team at the Memory Clinic make diagnoses according to clinical criteria, clinical AD diagnoses are only 'possible' or 'probable' in nature. A post-mortem examination of a patient's brain is required to confirm an AD diagnosis as 'definite' (McKhann et al., 1984). Studies comparing clinical and neuropathological diagnoses of AD have reported clinical diagnoses are accurate for 71% to 87% of cases (Beach, Monsell, Phillips, & Kukull, 2012; Joachim, Morris, & Selkoe, 1988; Lim et al., 1999; Wade et al., 1987), meaning that between 13% and 29% of cases with clinical diagnoses of AD are found to belong to other diagnostic categories. These data imply that a proportion of patients in the current AD group, may, in fact, have another condition, or a type of dementia other than AD.

Below, I outline two broad suggestions for the way forward regarding research and clinical use of the BNT-SA-SF, based on the study findings.

Change the BNT-SA-SF. The first broad suggestion is to change the BNT-SA-SF. There are a number of possible options here. Problematic items in the test could be replaced with (a) other items from the original test, or (b) with new items. On the other hand, more items could be added to the test to create a longer form, as longer versions are generally more discriminative than short versions in clinical populations (Tombaugh & Hubley, 1997). However, the discriminative capacity of the BNT-SA-SF appears reasonable, and the addition of more items would simply serve to increase administration time and to nullify the utility of having a short form that can be administered rapidly and that can be included in screening batteries easily.

I suggest that focus should be placed on substituting the last three items on the test, *compass*, *sphinx*, and *protractor*, due to the fact that so few participants in both the patient and control samples were able to name these items correctly. In addition, *compass* and *protractor* are from the same semantic category, both being tools used in geometry. With regards to face validity, it does not seem appropriate to have both items included in a test with only 15 items. If an individual does not know *compass* it is likely that they will not know *protractor* either. Therefore, it would perhaps be better to substitute at least one of these items to create more variation with items that reflect the wide range of semantic categories of the full 60-item test.

There are various ways in which new items could be selected. If replacement items were chosen based on the method in which items were originally selected for the BNT-SA-SF, *asparagus*, *latch*, or *tripod* could replace *compass*; *scroll*, *tongs*, or *yoke* could replace *sphinx*; and *trellis*, *palette*, or *abacus* could replace *protractor*. The BNT-SA-SF was

developed on an item-by-item bases, unlike many other short forms that are developed via odd-even or split-halves methods. Items were selected based on expert ratings of cultural appropriateness, picture ambiguity, difficulty, and colloquial use. A similar method of item selection has been used elsewhere in the creation of culturally modified BNT versions. For instance, in the creation of a widely-used Spanish adaptation, the Pontón-Satz BNT (Pontón et al., 1992), the authors selected 30 items from the original BNT using ratings from expert judges who ranked the items according to criteria such as cultural appropriateness, word frequency, and item difficulty. However, this qualitative approach to item selection has been critiqued in the literature. Roberts and Doucet (2011) suggest that authors cannot identify and replace problematic items based on intuition, and that the full test needs to be administered and problematic items then identified according to item difficulty, sensitivity, and name agreement (i.e., that there is only one possible correct name for the item). Some have approached item selection in this way, by using pilot studies and administering the test to a small group of participants in order to identify problematic items (Patricacou et al., 2007).

A better way to go about choosing items from the original test for inclusion in the short form, rather than intuition or expert opinion, is to use modern psychometric methods. One way to do this, which other authors have successfully used (e.g., Graves et al., 2004; Saxton, Ratcliff et al., 2000), would be to administer the full 60-item test to a large sample, representative of the population, and to then use item response theory (IRT) to identify items for inclusion in the final version of the test. IRT can be used to assess the differential functioning of the items, selecting those that perform equally across different demographic groups (Pedraza, Graff-Radford, et al., 2009). IRT can also be used to examine the psychometric properties of the test on an item-by-item level, selecting those items with the most desirable properties for inclusion in the short form (Pedraza, Sachs, Ferman, Rush, & Lucas, 2011). Examining the psychometric properties of the BNT using IRT, Pedraza et al. (2011) recommend this method for the development of BNT short forms or the replacement of problematic items to improve the psychometric properties of an existing test.

Realistically, it is unlikely that a single set of items will be culturally appropriate and function equally well across South Africa's diverse population, as Barker-Collo (2001) acknowledges with regard to the New Zealand population. However, it may be possible that the test can be made more appropriate while still maintaining or even improving its diagnostic validity and other psychometric properties. In addition, this method would also be beneficial in that it would preserve the existing advantages of the fact that BNT-SA-SF items are drawn from the pool of items from the original BNT, as discussed above.

Alternatively, if analysis fails to reveal any suitable replacements from the 60 items of the original test, alternative items can be sourced. This strategy would follow other modifications that have replaced certain BNT items with items that are specific to the language, culture, or geographical region in which the test is going to be used, often sourced from books, newspapers, or television (see, e.g., Patricacou et al., 2007). However, there are a number of considerations when choosing suitable items. For instance, in order to retain the properties of the original test, a replacement item should be of the same semantic category and of the same difficulty as the original item.

Keep the BNT-SA-SF. The second broad suggestion is to keep the BNT-SA-SF as it is at present and to norm this version more extensively, providing a large body of representative data from well-defined community-dwelling older individuals. Considering South Africa's linguistically diverse population, this would be a large but valuable undertaking, considering the lack of local normative data on psychological tests. Adopting such a strategy would reduce the potential for misdiagnosis due to the influence of demographic and sociocultural factors on test performance, and would, in the ideal scenario, provide clinicians and researchers with norms stratified by key demographic variables, such as age, education, and sex within different language groups. Such an approach would follow the suggestions of Shuttleworth-Jordan (1996), among others.

Even if one adopts this approach, however, it may still be necessary to modify the test slightly. Based on the fact that some of the items appear to be out of sequence, the order of the items would have to be rearranged to ensure that there is a clear linear trend according to respondent's naming accuracy, from easier items to those that are more difficult for the present study sample. Ordering the items in this manner follows the way the items are ordered in the original test and other translated or adapted versions (Mariën et al., 1998; Patricacou et al., 2007; Tallberg, 2005). Based on the item difficulty results for the control group, I rearranged the 15 BNT-SA-SF items in order of increasing difficulty (see Table G in Appendix G). It may be appropriate to use this order of items in future administrations of the BNT-SA-SF. Although it must be noted that as the general trend is of increasing difficulty as the test progresses, this exercise would not impact the validity of the test (Roberts et al., 2002). In addition, it is unclear whether the trend witnessed in this sample would be witnessed with other healthy population groups in South Africa.

In closing, it must be emphasized that the participants in this study were drawn from very specific demographic groups in the Western Cape region of South Africa, and that they are therefore not representative of the performance of the general South African population.

As noted in the first chapter, South Africa has 11 official languages and a culturally diverse population with varied experiential backgrounds, spread over nine provinces. BNT-SA-SF performance may vary substantially across different regional (e.g., urban vs. rural), linguistic, and cultural groups beyond those included in this study. For example, the population in the Western Cape has proportionately more Coloured (48.8%) than Black African (32.8%) and White (15.7%) people, whereas Black African people constitute the largest proportion of the population in the other eight provinces (Statistics South Africa, 2012). Similarly, whereas Afrikaans (49.7%) and English (20.2%) are two of the most frequently spoken languages in the Western Cape, isiZulu (22.7%) and isiXhosa (16.0%) are the most frequently spoken languages in South Africa as a whole (Statistics South Africa, 2012). Other African languages, including Sepedi, SiSwati, Setswana, and Sesotho are also predominant languages spoken in other provinces in the country. Furthermore, whereas in the Western Cape the proportion of the population with no education is low (2.7%), in other provinces this proportion is as high as 17.3%.

Future research on the BNT-SA-SF should look to include a much larger and more representative sample across both clinical and control participants. It would be particularly useful to find out whether the test is applicable to the other major cultural groups in South Africa, specifically Black African and Indian South African. Speaking generally about psychological test development in South Africa, Foxcroft (2004) states that tests should, at a minimum, be applicable to four main cultural groups in South Africa, namely Black African, Colored, Indian, and White. However, that the applicability of the BNT-SA-SF to two of these four groups (i.e., Black African and Indian) is not clear from the results of the present study does not undermine the utility of the current results for clinicians in the Western Cape who work with clients drawn from population groups similar to those in this study. In addition, the results of the present study may be useful to clinicians elsewhere in South Africa as a reference regarding the functioning of these BNT items and as a basis for using the BNT-SA-SF in assessing older adults who present with dementia-related referral questions.

Summary and Conclusion

In summary, the results suggest the BNT-SA-SF is a diagnostically valid instrument for distinguishing AD from normal aging. The test is sensitive to the naming deficits seen in AD, and patients with AD perform significantly more poorly than healthy controls. Further, that the BNT-SA-SF has good internal consistency, that the discriminative capacity of the test was not unduly affected by sociodemographic variables, and that the test appears sensitive to

dementia severity in AD (as measured by the MMSE), further supports its utility as a screening measure in identifying the naming impairment present in AD in South Africa.

The BNT-SA-SF did not distinguish AD from other types of dementia as well as it was able to distinguish AD from normal aging, however this study does provides some evidence that assessing confrontation naming using the BNT can discriminate other types of dementia from normal aging. Although the focus of this study was on the ability of the BNT-SA-SF to identify AD, results showed that the test may also be sensitive to naming deficits in the dementia subtypes included in the OD group, such as vascular dementia.

Overall, the results indicate that if an older adult (50-88 years) who is fluent in English, but for whom their first or home language is either English or Afrikaans, obtains a score on the BNT-SA-SF below the cut-scores identified in this study, this cannot be attributed readily to the individual's level of education, age, language, race, or SES. However, clinicians must consider the participants sex in interpreting his/her performance on the test. In addition, despite the diagnostic utility of the BNT-SA-SF, this study identified a number of items that may be problematic when used in South Africa, and one or more of the suggestions for the replacement of these items should be investigated in future studies.

To conclude, this study showed that an adapted BNT short form is a useful screening tool that can be used to help in the differential diagnosis of AD with English-fluent older adults in the Western Cape. This research is encouraging progress in the endeavor to provide locally appropriate and valid neuropsychological tests and representative normative data for use in South African clinical and research settings. It has practical value for clinicians and researchers working with older adult populations in South Africa, particularly in memory clinic settings or with those suspected of having dementia. The adapted test provides an alternate set of items for those that already use the BNT or any of the existing short forms in their clinical practice. It must be noted, however, that these results cannot necessarily be applied to individuals whose demographic profile differs from those in the study sample. The more generalized applicability of this modified test relies on the collection of more extensive data and further research investigating its validity and reliability in other South African population groups.

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., … Petersen, R. C. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*, 270–279.

Allegri, R. F., Villavicencio, A. F., Taragano, F. E., Rymberg, S., Mangone, C. A., & Baumann, D. (1997). Spanish Boston naming test norms. *The Clinical Neuropsychologist*, *11*, 416–420.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Andersen, K., Launer, L. J., Dewey, M. E., Letenneur, L., Ott, A., Copeland, J. R. M., … Hofman, A. (1999). Gender differences in the incidence of AD and vascular dementia: The EURODEM Studies. *Neurology*, *53*, 1992–1992.

Anderson, S. J. (2001). On the importance of collecting local neuropsychological normative data. *South African Journal of Psychology*, *31*, 29–35.

Armstrong, P., Lekezwa, B., & Siebrits, K. (2008). Poverty in South Africa: A profile based on recent household surveys. *The University of Stellenbosch: Stellenbosch economic working papers*, *4*.

Ashford, J. W., Kolm, P., Colliver, J. A., Bekian, C., & Hsu, L. (1989). Alzheimer patient evaluation and the mini-mental state: Item characteristic curve analysis. *Journal of Gerontology*, *44*, 139–146.

Au, R., Joung, P., Nicholas, M., Obler, L., Kass, R., & Albert, M. (1995). Naming ability across the adult life span. *Aging and Cognition*, *2*, 300–311.

Azrin, R. L., Mercury, M. G., Millsaps, D., Goldstein, D., Trejo, T., & Pliskin, N. H. (1996). Cautionary note on the Boston Naming Test: Cultural considerations (Abstract). *Archives of Clinical Neuropsychology*, *11*, 365–366.

Baillon, S., Muhommad, S., Marudkar, M., Suribhatla, S., Dennis, M., Spreadbury, C., … Lindesay, J. (2003). Neuropsychological performance in Alzheimer's disease and vascular dementia: Comparisons in a memory clinic population. *International Journal of Geriatric Psychiatry*, *18*, 602–608.

Balthazar, M. L. ., Cendes, F., & Damasceno, B. P. (2008). Semantic error patterns on the Boston Naming Test in normal aging, amnestic mild cognitive impairment, and mild Alzheimer's disease: Is there semantic disruption? *Neuropsychology*, *22*, 703–709.

Barbarotto, R., Capitani, E., Jori, T., Laiacona, M., & Molinari, S. (1998). Picture naming and progression of Alzheimer's disease: An analysis of error types. *Neuropsychologia*, *36*, 397–405.

Barker-Collo, S. (2001). The 60-Item Boston Naming Test: Cultural bias and possible adaptations for New Zealand. *Aphasiology*, *15*, 85–92.

Barker-Collo, S. (2007). Boston Naming Test performance of older New Zealand adults. *Aphasiology*, *21*, 1171–1180.

Barr, A., Benedict, R., Tune, L., & Brandt, J. (1992). Neuropsychological differentiation of Alzheimer's disease from vascular dementia. *International Journal of Geriatric Psychiatry*, *7*, 621–627.

Barratt, J., Khoza-Shangase, K., & Msimang, K. (2012). Speech-language assessment in a linguistically diverse setting: Preliminary exploration of the possible impact of informal "solutions" within the South African context. *South African Journal of Communication Disorders*, *59,* 34–44.

Barresi, B. A., Nicholas, M., Tabor Connor, L., Obler, L. K., & Albert, M. L. (2000). Semantic degradation and lexical access in age-related naming failures. *Aging, Neuropsychology, and Cognition*, *7*, 169–178.

Bayles, K. A., & Tomoeda, C. K. (1983). Confrontation naming impairment in dementia. *Brain and Language*, *19*, 98–114.

Beach, T. G., Monsell, S. E., Phillips, L. E., & Kukull, W. (2012). Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *Journal of Neuropathology*, *71*, 266–273.

Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society*, *11*, 215–227.

Bertolucci, P. H. F., Okamoto, I. H., Brucki, S. M. D., Siviero, M. O., Toniolo Neto, J., & Ramos, L. R. (2001). Applicability of the CERAD neuropsychological battery to Brazilian elderly. *Arquivos de Neuro-Psiquiatria*, *59*, 532–536.

Bialystok, E., & Craik, F. I. . (2007). Bilingualism and naming: Implications for cognitive assessment. *Journal of the International Neuropsychological Society*, *13*, 209–211.

Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism*, *10*, 210–223.

Brouillette, R. M., Martin, C. K., Correa, J. B., Davis, A. B., Han, H., Johnson, W. D., … Keller, J. N. (2011). Memory for Names test provides a useful confrontational naming task for aging and continuum of dementia. *Journal of Alzheimer's Disease*, *23*, 665–671.

Bucks, R. S., Ashworth, D. L., Wilcock, G. K., & Siegfried, K. (1996). Assessment of activities of daily living in dementia: Development of the Bristol Activities of Daily Living Scale. *Age and Aging*, *25*, 113 – 120.

Burke, W. J., Roccaforte, W. H., & Wengel, S. P. (1991). The Short Form of the Geriatric Depression Scale: A Comparison With the 30-Item Form. *Journal of Geriatric Psychiatry and Neurology*, *4*, 173–178.

Busch, R. M., Frazier, T. W., Haggerty, K. A., & Kubu, C. S. (2005). Utility of the Boston Naming Test in predicting ultimate side of surgery in patients with medically intractable temporal lobe epilepsy. *Epilepsia*, *46*, 1773–1779.

Calero, M. D, Arnedo, M. L., Navarro, E., Ruiz-Pedrosa, M., & Carnero, C. (2002). Usefulness of a 15-item version of the Boston Naming Test in neuropsychological assessment of low-educational elders with dementia. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*, 187–191.

Chen, P., Ratcliff, G., Belle, S. H., Cauley, J. A., DeKosky, S. T., & Ganguli, M. (2001). Patterns of cognitive decline in presymptomatic Alzheimer disease: A prospective community study. *Archives of general psychiatry*, *58*, 853–858.

Chertkow, H., & Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type what do various measures measure? *Brain*, *113*, 397–417.

Chertkow, H., Bub, D., & Seidenberg, M. (1989). Priming and semantic memory loss in Alzheimer's disease. *Brain and language*, *36*, 420–446.

Cheung, R. W., Cheung, M. C., & Chan, A. S. (2004). Confrontation naming in Chinese patients with left, right or bilateral brain damage. *Journal of the International Neuropsychological Society*, *10*, 46–53.

Chosak Reiter, J. (2000). Measuring cognitive processes underlying picture naming in Alzheimer's and cerebrovascular dementia: A general processing tree approach. *Journal of clinical and experimental Neuropsychology*, *22*, 351–369.

Connor, L. T., Spiro, A., Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *59*, 203–209.

Cruice, M. N., Worrall, L. E., & Hickson, L. M. H. (2000). Boston Naming Test results for healthy older Australians: A longitudinal and cross-sectional study. *Aphasiology*, *14*, 143–155.

D'ath, P., Katona, P., Mullan, E., Evans, S., & Katona, C. (1994). Screening, detection and management of depression in elderly primary care attenders. I: The acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. *Family Practice*, *11*, 260–266.

Davies, K. G., Bell, B. D., Bush, A. J., Hermann, B. P., Dohan, C., & Jaap, A. S. (1998). Naming Decline after left anterior temporal lobectomy correlates with pathological status of resected hippocampus. *Epilepsia*, *39*, 407–419.

De Jager, C. A., Hogervorst, E., Combrinck, M., & Budge, M. M. (2003). Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological medicine*, *33*, 1039–1050.

De la Plata, C. M., Arango-Lasprilla, J. C., Alegret, M., Moreno, A., Tárraga, L., Lara, M., … Cullum, C. M. (2009). Item analysis of three Spanish naming tests: A cross-cultural investigation. *NeuroRehabilitation*, *24*, 75–85.

De la Torre, J. C. (2004). Is Alzheimer's disease a neurodegenerative or a vascular disorder? Data, dogma, and dialectics. *The Lancet Neurology*, *3*, 184–190.

Del Toro, C. M., Bislick, L. P., Comer, M., Velozo, C., Romero, S., Gonzalez Rothi, L. J., & Kendall, D. L. (2011). Development of a short form of the Boston Naming Test for individuals with aphasia. *Journal of Speech, Language, and Hearing Research*, *54*, 1089–1100.

DeLeon, J., Gottesman, R. F., Kleinman, J. T., Newhart, M., Davis, C., Heidler-Gary, J., … Hillis, A. E. (2007). Neural regions essential for distinct cognitive processes underlying picture naming. *Brain*, *130*, 1408–1422.

Diehl, J., Monsch, A. U., Aebi, C., Wagenpfeil, S., Krapp, S., Grimmer, T., … Kurz, A. (2005). Frontotemporal dementia, semantic dementia, and Alzheimer's disease: The contribution of standard neuropsychological tests to differential diagnosis. *Journal of Geriatric Psychiatry & Neurology*, *18*, 39–44.

Dotson, V. M., Kitner-Triolo, M. H., Evans, M. K., & Zonderman, A. B. (2009). Effects of race and socioeconomic status on the relative influence of education and literacy on cognitive functioning. *Journal of the International Neuropsychological Society*, *15*, 580.

Faber-Langendoen, K., Morris, J. C., Knesevich, J. W., LaBarge, E., Miller, J. P., & Berg, L. (1988). Aphasia in senile dementia of the Alzheimer type. *Annals of Neurology*, *23*, 365–370.

Farmer, A. (1990). Performance of normal males on the Boston Naming Test and the Word Test. *Aphasiology*, *4*, 293–296.

Fastenau, P. (1998). Validity of regression-based norms: An empirical test of the comprehensive norms with older adults. *Journal of clinical and experimental neuropsychology*, *20*, 906–916.

Fastenau, P., Denburg, N., & Mauer, B. (1998). Parallel short forms for the Boston Naming Test: Psychometric properties and norms for older adults. *Journal of Clinical and Experimental Neuropsychology*, *20*, 828–834.

Ferman, T. J., Smith, G. E., Boeve, B. F., Graff-Radford, N. R., Lucas, J. A., Knopman, D. S., … Dickson, D. W. (2006). Neuropsychological differentiation of dementia with Lewy Bodies from normal aging and Alzheimer's disease. *The Clinical Neuropsychologist*, *20*, 623–636.

Ferraro, F. R., & Barth, J. (2003). Speeded lexical decision performance on 15-item forms of the Boston Naming Test. *Psychology and Education*, *40*, 38–40.

Ferrett, H. L., Carey, P. D., Thomas, K. G. F., Tapert, S. F., & Fein, G. (2010). Neuropsychological performance of South African treatment-naïve adolescents with alcohol dependence. *Drug and Alcohol Dependence*, *110*, 8–14.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, California: Sage Publications Ltd.

Fillenbaum, G. G., Huber, M., & Taussig, I. M. (1997). Performance of elderly White and African American community residents on the abbreviated CERAD Boston Naming Test. *Journal of clinical and experimental neuropsychology*, *19*, 204–210.

Fisher, N. J., Tierney, M. C., Snow, W. G., & Szalai, J. P. (1999). Odd/even short forms of the Boston Naming Test: Preliminary geriatric norms. *The Clinical Neuropsychologist*, *13*, 359–364.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Foxcroft, C. D. (2004). Planning a psychological test in the multicultural South African context. *South African Journal of Industrial Psychology*, *30*, 8–15.

Franzen, M. D., Haut, M. W., Rankin, E., & Keefover, R. (1995). Empirical comparison of alternate forms of the Boston Naming Test. *The Clinical Neuropsychologist*, *9*, 225–229.

Gainotti, G. (1987). The status of the semantic-lexical structures in anomia. *Aphasiology*, *1*, 449–461.

Gao, S., Hendrie, H. C., Hall, S. K., & Hui, S. (1998). The relationships between age, sex, and the incidence of dementia and Alzheimer disease: A meta-analysis. *Archives of General Psychiatry*, *55*, 809–815.

Garrard, P., Lambon Ralph, M. A., Patterson, K., Pratt, K. H., & Hodges, J. R. (2005). Semantic feature knowledge and picture naming in dementia of Alzheimer's type: A new approach. *Brain and language*, *93*, 79–94.

Gauthier, S., Gélinas, I., & Gauthier, L. (1997). Functional disability in Alzheimer's disease. *International psychogeriatrics*, *9*, 163–165.

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*, 61–105.

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society*, *13*, 197–208.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The Boston Diagnostic Aphasia Examination* (3rd ed.). Philadelphia: Lippincott Williams and Wilkins.

Goodglass, H., Kaplan, E., & Weintraub, S. (1976). Boston Naming Test (Experimental ed.). Boston: Boston University.

Goulet, P., & Ska, B. (1994). Is there a decline in picture naming with advancing age? *Journal of Speech, Language and Hearing Research*, *37*, 629.

Graves, R. E., Bezeau, S. C., Fogarty, J., & Blair, R. (2004). Boston Naming Test short forms: A comparison of previous forms with new item response theory based forms. *Journal of Clinical and Experimental Neuropsychology*, *26*, 891–902.

Grieve, K. W., & Cave, J. (2009). The effect of quality of education on neuropsychological test performance. *New Voices in Psychology*, *5*, 29–48.

Grieve, K. W., & Van Eeden, R. (2010). A preliminary investigation of the suitability of the WAIS-III for Afrikaans-speaking South Africans. *South African Journal of Psychology, 40,* 262-271.

Grigoletto, F., Zappalà, G., Anderson, D. W., & Lebowitz, B. D. (1999). Norms for the Mini-Mental State Examination in a healthy population. *Neurology*, *53*, 315–315.

Grossman, M., McMillan, C., Moore, P., Ding, L., Glosser, G., Work, M., & Gee, J. (2004). What's in a name: Voxel- based morphometric analyses of MRI and naming difficulty in Alzheimer's disease, frontotemporal dementia and corticobasal degeneration. *Brain*, *127*, 628–649.

Hall, J. R., Vo, H. T., Johnson, L. A., Wiechmann, A., & O'Bryant, S. E. (2012). Boston Naming Test: Gender differences in older adults with and without Alzheimer's dementia. *Psychology*, *3*, 485–488.

Hawkins, K. A., & Bender, S. (2002). Norms and the relationship of Boston Naming Test performance to vocabulary and education: A review. *Aphasiology*, *16*, 1143–1153.

Hawkins, K. A., Cromer, J. R., Piotrowski, A. S., & Pearlson, G. D. (2011). Mini-Mental State Exam performance of older African Americans: Effect of age, gender, education, hypertension, diabetes, and the inclusion of serial 7s subtraction versus "world" backward on score. *Archives of Clinical Neuropsychology*, *26*, 645–652.

Hawkins, K. A., Sledge, W. H., Orleans, J. F., Quinlan, D. M., Rakfeldt, J., & Huffman, R. E. (1993). Normative implications of the relationship between reading vocabulary and Boston Naming Test performance. *Archives of Clinical Neuropsychology*, *8*, 525–537.

Heaton, R., Avitable, N., Grant, I., & Matthews, C. (1999). Further cross validation of regression-based neuropsychological norms with an update for the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology*, *21*, 572–182.

Henderson, V. W., Mack, W., Freed, D. M., Kempler, D., & Andersen, E. S. (1990). Naming consistency in Alzheimer's disease. *Brain and language*, *39*, 530–538.

Herman, A. A., Stein, D. J., Seedat, S., Heeringa, S. G., Moomal, H., & Williams, D. R. (2009). The South African Stress and Health (SASH) study: 12-month and lifetime prevalence of common mental disorders. *South African Medical Journal*, *99*, 339–344.

Herrmann, N., Mittmann, N., Silver, I. L., Shulman, K. I., Busto, U. A., Shear, N. H., & Naranjo, C. A. (1996). A validation study of The Geriatric Depression Scale short form. *International Journal of Geriatric Psychiatry*, *11*, 457–460.

Hobson, V. L., Hall, J. R., Harvey, M., Cullum, C. M., Lacritz, L., Massman, P. J., … O'Bryant, S. E. (2011). An examination of the Boston Naming Test: Calculation of "estimated" 60-item score from 30- and 15-item scores in a cognitively impaired population. *International Journal of Geriatric Psychiatry*, *26*, 351–355.

Hodges, J. R., Salmon, D. P., & Butters, N. (1991). The nature of the naming deficit in Alzheimer's and Huntington's disease. *Brain, 114*, 1547–1558.

Hodges, J. R., Salmon, D. P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia*, *30*, 301–314.

Huff, F. J. (1986). Equivalent forms of the Boston naming test. *Journal of clinical and experimental neuropsychology*, *8*, 556–562.

Huff, F. J., Corkin, S., & Growdon, J. H. (1986). Semantic impairment and anomia in Alzheimer's disease. *Brain and Language*, *28*, 235–249.

Jacobs, D. M., Sano, M., Dooneief, G., Marder, K., Bell, K. L., & Stern, Y. (1995). Neuropsychological detection and characterization of preclinical Alzheimer's disease. *Neurology*, *45*, 957–962.

Jefferson, A., Wong, S., Gracer, T., Ozonoff, A., Green, R., & Stern, R. (2007). Geriatric performance on an abbreviated version of the Boston Naming Test. *Applied Neuropsychology*, *14*, 215–223.

Joachim, C. L., Morris, J. H., & Selkoe, D. J. (1988). Clinically diagnosed Alzheimer's disease: Autopsy results in 150 cases. *Annals of neurology*, *24*, 50–56.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, *120*, 113–139.

Joubert, S., Brambati, S. M., Ansado, J., Barbeau, E. J., Felician, O., Didic, M., … Kergoat, M.-J. (2010). The cognitive and neural expression of semantic memory impairment in

mild cognitive impairment and early Alzheimer's disease. *Neuropsychologia*, *48*, 978–988.

Kalula, S. Z., Ferreira, M., Thomas, K. G. ., de Villiers, L., Joska, J. A., & Geffen, L. N. (2010). Profile and management of patients at a memory clinic. *South African Medical Journal*, *100*, 449–451.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia: Lea & Febiger.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

Katz, M. J., Lipton, R. B., Hall, C. B., Zimmerman, M. E., Sanders, A. E., Verghese, J., … Derby, C. A. (2012). Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and Alzheimer dementia in blacks and whites: A report from the Einstein Aging Study. *Alzheimer Disease & Associated Disorders*, *26*, 335–343.

Kavé, G. (2005). Standardization and norms for a Hebrew naming test. *Brain and Language*, *92*, 204–211.

Kendall, D. L., Rodriguez, A. D., Rosenbek, J. C., Conway, T., & Rothi, L. J. G. (2006). Influence of intensive phonomotor rehabilitation on apraxia of speech. *Journal of rehabilitation research and development*, *43*, 409–418.

Kent, P. S., & Luszcz, M. A. (2002). A Review of the Boston Naming Test and multiple-occasion normative data for older adults on 15-item versions. *The Clinical Neuropsychologist*, *16*, 555–575.

Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment, 5,* 395–399.

Killgore, W. D., & Adams, R. L. (1999). Prediction of Boston Naming Test performance from vocabulary scores: Preliminary guidelines for interpretation. *Perceptual and Motor Skills*, *89*, 327–337.

Kim, H., Kim, E. Y., & Na, D. L. (1997). Naming deficits in patients with dementia of the Alzheimer type: Error analysis of Korean version-Boston Naming Test. *Journal of the Korean Neurological Association*, *15*, 1012–1021.

Kim, H., & Na, D. L. (1999). Normative data on the Korean version of the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology*, *21*, 127–134.

Kimbarow, M. L., Vangel, S. J., & Lichtenberg, P. A. (1996). The influence of demographic variables on normal elderly subjects' performance on the Boston Naming Test. *Clinical Aphasiology*, *24*, 135–144.

Kirshner, H. S., Webb, W. G., & Kelly, M. P. (1984). The naming disorder of dementia. *Neuropsychologia*, *22*, 23–30.

Knesevich, J. W., LaBarge, E., & Edwards, D. (1986). Predictive value of the Boston Naming Test in mild senile dementia of the Alzheimer type. *Psychiatry research*, *19*, 155–161.

Kohn, S. E., & Goodglass, H. (1985). Picture-naming in aphasia. *Brain and Language, 24*, 266–283.

Kohnert, K. J., Hernandez, A. E., & Bates, E. (1998). Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain and language*, *65*, 422–440.

Kramer, J. H., Jurik, J., Sharon, J. S., Rankin, K. P., Rosen, H. J., Johnson, J. K., & Miller, B. L. (2003). Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease. *Cognitive and Behavioral Neurology, 16*, 211–218.

LaBarge, E., Balota, D. A., Storandt, M., & Smith, D. S. (1992). An analysis of confrontation naming errors in senile dementia of the Alzheimer type. *Neuropsychology*, *6*, 77.

LaBarge, E., Edwards, D., & Knesevich, J. W. (1986). Performance of normal elderly on the Boston Naming Test. *Brain and Language, 27*, 380–384.

Laine, M., Goodglass, H., Niemi, J., Koivuselka-Sallinen, P., Tuomainen, J., & Marttila, R. (1993). Adaptation of the Boston Diagnostic Aphasia Examination and the Boston Naming Test into Finnish. *Logopedics Phonatrics Vocology*, *18*, 83–92.

Lancu, I., & Olmer, A. (2006). The Mini-Mental State Examination: An up-to-date review. *Harefuah*, *145*, 687–690.

Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Randolph, C. (1999). An empirically derived short form of the Boston Naming Test. *Archives of Clinical Neuropsychology*, *14*, 481–487.

Larrain, C. M., & Cimino, C. R. (1998). Alternate forms of the Boston Naming Test in Alzheimer's Disease. *Clinical Neuropsychologist*, *12*, 525–530.

Le Dorze, G., & Durocher, J. (1992). The effects of age, educational level, and stimulus length on naming in normal subjects. *Journal of Speech-Language Pathology and Audiology*, *16*, 21–29.

Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*, 223–232.

Lezak, M. D., Whitman, R., & Bourdette, D. (1990). Emotional impact of cognitive inefficiencies in multiple sclerosis (Abstract). *Journal of Clinical and Experimental Neuropsychology*, *8*, 50.

Lezak, M. D., Howieson, D., & Loring, D. (2004). *Neuropsychological assessment* (4th ed.). Oxford: Oxford University Press.

Lim, A., Tsuang, D., Kukull, W., Nochlin, D., Leverenz, J., McCormick, W., … Peskind, E. R. (1999). Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series. *Journal of the American Geriatrics Society*, *47*, 564–569.

Loring, D. W., Strauss, E., Hermann, B. P., Barr, W. B., Perrine, K., Trenerry, M. R., … Meador, K. J. (2008). Differential neuropsychological test sensitivity to left temporal lobe epilepsy. *Journal of the International Neuropsychological Society*, *14*, 394–400.

Lucas, J. A., Ivnik, R. J., Smith, G. E., Ferman, T. J., Willis, F. B., Petersen, R. C., & Graff-Radford, N. R. (2005). Mayo's older African Americans normative studies: Norms for Boston Naming Test, Controlled Oral Word Association, Category Fluency, Animal Naming, Token Test, Wrat-3 Reading, Trail Making Test, Stroop Test, and Judgment of Line Orientation. *Clinical Neuropsychologist*, *19*, 243–269.

Lukatela, K., Malloy, P., Jenkins, M., & Cohen, R. (1998). The naming deficit in early Alzheimer's and vascular dementia. *Neuropsychology*, *12*, 565–572.

MacGinite, W., & MacGinite, R. K. (1989). *Gates-MacGinite Reading Test: Manual for scoring and interpretation, levels 7/9 and 10/12, forms K and L* (3rd ed.). Chicago, IL: Riverside Publishing.

Mack, W. J., Freed, D. M., Williams, B. W., & Henderson, V. W. (1992). Boston Naming Test: Shortened versions for use in Alzheimer's disease. *The Journal of Gerontology*, *47*, 154–158.

MacKay, A., Connor, L. T., & Storandt, M. (2005). Dementia does not explain correlation between age and scores on the Boston Naming Test. *Archives of Clinical Neuropsychology*, *20*, 129–133.

Manly, J. J. (2005). Advantages and disadvantages of separate norms for African Americans. *The Clinical Neuropsychologist*, *19*, 270–275.

Manly, J. J., Byrd, D. A., Touradji, P., & Stern, Y. (2004). Acculturation, reading level, and neuropsychological test performance among African American elders. *Applied Neuropsychology*, *11*, 37–46.

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, *8*, 341–348.

Marc, L. G., Raue, P. J., & Bruce, M. L. (2008). Screening performance of the Geriatric Depression Scale (GDS-15) in a diverse elderly home care population. *The American Journal of Geriatric Psychiatry*, *16*, 914–921.

Mariën, P., Mampaey, E., Vervaet, A., Saerens, J., & De Deyn, P. P. (1998). Normative data for the Boston Naming Test in native Dutch-speaking Belgian elderly. *Brain and language*, *65*, 447–467.

Martin, A., & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and language*, *19*, 124–141.

Mathias, J. L., & Burke, J. (2009). Cognitive functioning in Alzheimer's and vascular dementia: a meta-analysis. *Neuropsychology*, *23*, 411–423.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*, *34*, 939–939.

Mendez, M. F., Cherrier, M., Perryman, K. M., Pachana, N., Miller, B. L., & Cummings, J. L. (1996). Frontotemporal dementia versus Alzheimer's disease: Differential cognitive features. *Neurology*, *47*, 1189–1194.

Mendoça, J. (2010). *Performance of English, Zulu and Sotho students on the Boston Naming Test: An Investigation into the items responsible for cultural bias* (Unpublished masters dissertation). University of the Witwatersrand, Johannesburg, South Africa.

Miller, K. M., Finney, G. R., Meador, K. J., & Loring, D. W. (2010). Auditory responsive naming versus visual confrontation naming in dementia. *The Clinical Neuropsychologist*, *24*, 103–118.

Miotto, E. C., Sato, J., Lucia, M., Camargo, C. H., & Scaff, M. (2010). Development of an adapted version of the Boston Naming Test for Portuguese speakers. *Revista Brasileira de Psiquiatria*, *32*, 279–282.

Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, *43*, 411–431.

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Mitrushina, M., & Satz, P. (1989). Differential decline of specific memory components in normal aging. *Brain Dysfunction, 2,* 330–335.

Mitrushina, M., & Satz, P. (1995). Repeated testing of normal elderly with the Boston Naming Test. *Aging*, *7,* 123–127.

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology, 43,* 2412–2414.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., … Clark, C. (1989). The Consortium to Establish a Registry For Alzheimer's Disease (CERAD): I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39*, 1159–1165.

Morris, J. C., Storandt, M., Miller, J. P., McKeel, D. W., Price, J. L., Rubin, E. H., & Berg, L. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of neurology*, *58*, 397.

Mosdell, J., Balchin, R., & Ameen, O. (2010). Adaptation of aphasia tests for neurocognitive screening in South Africa. *South African Journal of Psychology*, *40*, 250–261.

Myer, L., Stein, D. J., Grimsrud, A., Seedat, S., & Williams, D. R. (2008). Social determinants of psychological distress in a nationally-representative sample of South African adults. *Social science & medicine*, *66*, 1828–1840.

Nebes, R. D. (1989). Semantic memory in Alzheimer's disease. *Psychological bulletin*, *106*, 377–394.

Nebreda, M. C., García-Caballero, A., Asensio, E., Revilla, P., Rodriguez-Girondo, M., & Mateos, R. (2011). A short-form version of the Boston Naming Test for language screening in dementia in a bilingual rural community in Galicia (Spain). *International Psychogeriatrics*, *23*, 435–441.

Neils, J., Baris, J. M., Carter, C., Dell'aira, A. L., Nordloh, S. J., Weiler, E., & Weisiger, B. (1995). Effects of age, education, and living environment on Boston Naming Test performance. *Journal of Speech, Language and Hearing Research*, *38*, 1143–1149.

Neils, J., Brennan, M. M., Cole, M., Boller, F., & Gerdeman, B. (1988). The use of phonemic cueing with Alzheimer's disease patients. *Neuropsychologia*, *26*, 351–354.

Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practice*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Nicholas, L. E., Brookshire, R. H., Maclennan, D. L., Schumacher, J. G., & Porrazzo, S. A. (1989). Revised administration and scoring procedures for the Boston Naming Test and norms for non-brain-damaged adults. *Aphasiology*, *3*, 569–580.

Nicholas, M., Obler, L., Albert, M., & Goodglass, H. (1985). Lexical retrieval in heathy aging. *Cortex*, *21*, 595–606.

Nield, M. (2007). Using neuropsychological test performance to identify Alzheimer's disease at a South African memory clinic (Unpublished thesis). University of Cape Town, Cape Town, South Africa.

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental science*, *10*, 464–480.

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science*, *8*, 74–87.

Noe, E., Marder, K., Bell, K. L., Jacobs, D. M., Manly, J. J., & Stern, Y. (2004). Comparison of dementia with Lewy bodies to Alzheimer's disease and Parkinson's disease with dementia. *Movement Disorders*, *19*, 60–67.

Owen, K. (1992). The suitability of Raven's standard progressive matrices for various groups in South Africa. *Personality and Individual Differences*, *13*, 149–159.

Pachana, N. A., Boone, K. B., Miller, B. L., Cummings, J. L., & Berman, N. (1996). Comparison on neuropsychological functioning in Alzheimer's disease and frontotemporal dementia. *Journal of the International Neuropsychological Society*, *2*, 505–510.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255.

Paivio, A., Clark, J. M., Digdon, N., & Bons, T. (1989). Referential processing: Reciprocity and correlates of naming and imaging. *Memory & Cognition*, *17*, 163–174.

Patricacou, A., Psallida, E., Pring, T., & Dipper, L. (2007). The Boston Naming Test in Greek: Normative data and the effects of age and education on naming. *Aphasiology*, *21*, 1157–1170.

Pedraza, O., Dotson, V. M., Willis, F. B., Graff-Radford, N. R., & Lucas, J. A. (2009). Internal consistency and test-retest stability of the Geriatric Depression Scale-Short Form in African American older adults. *Journal of psychopathology and behavioral assessment*, *31*, 412–416.

Pedraza, O., Graff-Radford, N., Smith, G., Ivnik, R., Willis, F. B., Petersen, R., & Lucas, J. A. (2009). Differential item functioning of the Boston Naming Test in cognitively normal African American and Caucasian older adults. *Journal of the International Neuropsychological Society*, *15*, 758–768.

Pedraza, O., Sachs, B. C., Ferman, T. J., Rush, B. K., & Lucas, J. A. (2011). Difficulty and discrimination parameters of Boston Naming Test items in a consecutive clinical series. *Archives of Clinical Neuropsychology, 26,* 434–444.

Peña-Casanova, J., Quiñones-Úbeda, S., Gramunt-Fombuena, N., Aguilar, M., Casas, L., Molinuevo, J. L., … Antúnez, C. (2009). Spanish Multicenter Normative Studies (NEURONORMA project): Norms for Boston naming test and token test. *Archives of Clinical Neuropsychology*, *24*, 343–354.

Pepe, M.S., Janes, H., Longton, G., Leisenring, W., Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology, 159,* 882–890.

Perneczky, R., Wagenpfeil, S., Komossa, K., Grimmer, T., Diehl, J., & Kurz, A. (2006). Mapping scores onto stages: Mini-Mental State Examination and Clinical Dementia Rating. *American Journal of Geriatric Psych*, *14*, 139–144.

Piguet, O., Millar, J. L., Bennett, H. P., Lye, T. C., Creasey, H., & Broe, G. A. (2001). Boston Naming Test: Normative data for older Australians. *Brain Impairment*, *2*, 131–139.

Pontón, M. O., Satz, P., Herrera, L., Young, R., Ortiz, F., D'Elia, L., … Namerow, N. (1992). Modified Spanish version of the Boston Naming Test. *Clinical Neuropsychologist*, *3*, 334.

Posel, D. (2011). Adult literacy rates in South Africa: A comparison of different measures. *Language Matters*, *42*, 39–49.

Price, B. H., Gurvit, H., Weintraub, S., Geula, C., Leimkuhler, E., & Mesulam, M. (1993). Neuropsychological patterns and language deficits in 20 consecutive cases of autopsy-confirmed Alzheimer's disease. *Archives of Neurology*, *50*, 931–937.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*, 33–65.

Randolph, C., Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Hermann, B. P. (1999). Determinants of confrontation naming performance. *Archives of Clinical Neuropsychology*, *14*, 489–496.

Robbins, R. N., Joska, J. A., Thomas, K. G. F., Stein, D. J., Linda, T., Mellins, C. A., & Remien, R. H. (2013). Exploring the utility of the Montreal Cognitive Assessment to detect HIV-associated neurocognitive disorder: The challenge and need for culturally valid screening tests in South Africa. *The Clinical Neuropsychologist*, *27*, 437–454.

Roberts, P. M., & Doucet, N. (2011). Performance of French-speaking Quebec adults on the Boston Naming Test. *Canadian Journal of Speech-Language Pathology and Audiology, 35,* 254–267.

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, *16*, 635–645.

Rochford, G. (1971). A study of naming errors in dysphasic and in demented patients. *Neuropsychologia*, *9*, 437–443.

Rogers, S. L., & Friedman, R. B. (2008). The underlying mechanisms of semantic memory loss in Alzheimer's disease and semantic dementia. *Neuropsychologia*, *46*, 12–21.

Rogers, T.T., Ivanoiu, A., Patterson, K., & Hodges, J. R. (2006). Semantic memory in Alzheimer's disease and the frontotemporal dementias: A longitudinal study of 236 patients. *Neuropsychology*, *20*, 319–335.

Roos, A., Calata, D., Jonkers, L., Maritz, S. J., Kidd, M., Daniels, W. M., & Hugo, F. J. (2010). Normative data for the Tygerberg Cognitive Battery and Mini-Mental Status Examination in a South African population. *Comprehensive Psychiatry*, *51*, 207–216.

Ross, T. P., Lichtenberg, P. A., & Christensen, B. K. (1995). Normative data on the Boston Naming Test for elderly adults in a demographically diverse medical sample. *The Clinical Neuropsychologist*, *9*, 321–325.

Ruitenberg, A., Ott, A., van Swieten, J. C., Hofman, A., & Breteler, M. (2001). Incidence of dementia: Does gender make a difference? *Neurobiology of Aging, 22*, 575–580.

Saxton, J., Munro, C. A., Butters, M. A., Schramke, C., & McNeil, M. A. (2000). Alcohol, dementia, and Alzheimer's disease: Comparison of neuropsychological profiles. *Journal Of Geriatric Psychiatry And Neurology, 13*, 141–149.

Saxton, J., Ratcliff, G., Munro, C. A., Coffey, E. C., Becker, J. T., Fried, L., & Kuller, L. (2000). Normative Data on the Boston Naming Test and two equivalent 30-item short forms. *Clinical Neuropsychologist, 14*, 526–534.

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development, 61*, 1728–1743.

Schmidtke, K., & Hüll, M. (2002). Neuropsychological differentiation of small vessel disease, Alzheimer's disease and mixed dementia. *Journal of the Neurological sciences, 203*, 17–22.

Schmitter-Edgecombe, M., Vesneski, M., & Jones, D. W. R. (2000). Aging and word-finding: A comparison of spontaneous and strained naming tests. *Archives of Clinical Neuropsychology, 15*, 479–493.

Serrano, C., Allegri, R. F., Drake, M., Butman, J., Harris, P., Nagle, C., & Ranalli, C. (2001). A shortened form of the Spanish Boston naming test: A useful tool for the diagnosis of Alzheimer's disease (Abstract). *Revista de Neurologia, 33*, 624.

Seymour, P. H. K. (1973). A model for reading, naming and comparison. *British Journal of Psychology, 64*, 35–49.

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. In T. Brink (Ed.), *Clinical gerontology: A guide to assessment and intervention* (pp.165-173). New York: Hayworth Press.

Shuttleworth, E. C., & Huber, S. J. (1988). The naming disorder of dementia of Alzheimer type. *Brain and Language, 34*, 222–234.

Shuttleworth-Edwards, A. B., Kemp, R. D., Rust, A. L., Muirhead, J. G. L., Hartman, N. P., & Radloff, S. E. (2004). Cross-cultural effects on IQ test performance: A review and preliminary normative indications. *Journal of Clinical and Experimental Neuropsychology, 26*, 903–920.

Shuttleworth-Edwards, A. B., Martin, J. R., Reid, I., & Radloff, S.E. (2004). A cross-cultural study with culture fair normative indications on WAIS-III digit symbol -incidental leaning. *Journal of Clinical and Experimental Neuropsychology, 26,* 921–932.

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*, 96–102.

Sikkes, S. A. M., De Lange-de Klerk, E. S. M., Pijnenburg, Y. A. L., & Scheltens, P. (2009). A systematic review of Instrumental Activities of Daily Living scales in dementia: Room for improvement. *Journal of Neurology, Neurosurgery & Psychiatry*, *80*, 7–12.

Simos, P. G., Kasselimis, D., & Mouzaki, A. (2011). Age, gender, and education effects on vocabulary measures in Greek. *Aphasiology*, *25*, 475–491.

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*, *30*, 1413–1425.

Statistics South Africa. (2012). *Census 2011: Census in brief*. Pretoria: Statistics South Africa.

Stewart, R., Richards, M., Brayne, C., & Mann, A. (2001). Cognitive function in UK community-dwelling African Caribbean elders: Normative data for a test battery. *International journal of geriatric psychiatry*, *16*, 518–527.

Storms, G., Saerens, J., & De Deyn, P. P. (2004). Normative data for the Boston Naming Test in native Dutch-speaking Belgian children and the relation with intelligence. *Brain and language*, *91*, 274–281.

Strauss, E., Sherman, E. M. ., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1–26.

Tallberg, I. M. (2005). The Boston Naming Test in Swedish: Normative data. *Brain and Language*, *94*, 19–31.

Taussig, I. M., Henderson, V. W., & Mack, W. (1992). Spanish translation and validation of a neuropsychological battery: Performance of Spanish-and English-speaking Alzheimer's disease patients and normal comparison subjects. *Clinical Gerontologist*, *11*, 95–108.

Testa, J. A., Ivnik, R. J., Boeve, B., Petersen, R. C., Pankratz, V. S., Knopman, D., … Smith, G. E. (2004). Confrontation naming does not add incremental diagnostic utility in MCI and Alzheimer's disease. *Journal of the International Neuropsychological Society*, *10*, 504–512.

Thompson, L. L., & Heaton, R. K. (1989). Comparison of different versions of the Boston Naming Test. *The Clinical Neuropsychologist*, *3*, 184–192.

Tombaugh, T., & Hubley, A. (1997). The 60-item Boston Naming Test: Norms for cognitively intact adults aged 25 to 88 years. *Journal of Clinical and Experimental Neuropsychology*, *19*, 922–932.

Tombaugh, T., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society, 40,* 922–935.

Touradji, P., Manly, J. J., Jacobs, D. M., & Stern, Y. (2001). Neuropsychological test performance: A study of Non-Hispanic white elderly. *Journal of Clinical & Experimental Neuropsychology*, *23*, 643 – 649.

Van Gorp, W. G., Satz, P., Kiersch, M. E., & Henry, R. (1986). Normative data on the Boston Maming Test for a group of normal older adults. *Journal of Clinical and Experimental Neuropsychology*, *8*, 702–705.

Van Schalkwyk, G., Botha, H., & Seedat, S. (2012). Comparison of 2 dementia screeners, the Test Your Memory Test and the Mini-Mental State Examination, in a primary care setting. *Journal of Geriatric Psychiatry and Neurology*, *25*, 85–88.

Villardita, C. (1993). Alzheimer's disease compared with cerebrovascular dementia: Neuropsychological similarities and differences. *Acta Neurologica Scandinavica*, *87*, 299–308.

Wade, J. P., Mirsen, T. R., Hachinski, V. C., Fisman, M., Lau, C., & Merskey, H. (1987). The clinical diagnosis of Alzheimer's disease. *Archives of neurology*, *44*, 24–29.

Wancata, J., Alexandrowicz, R., Marquart, B., Weiss, M., & Friedrich, F. (2006). The criterion validity of the Geriatric Depression Scale: A systematic review. *Acta Psychiatrica Scandinavica*, *114*, 398–410.

Watson, M., McMahon, M., Foxcroft, C., & Els, C. (2010). Occupational aspirations of low socioeconomic black South African children. *Journal of Career Development*, *37*, 717–734.

Wechsler, D. (1981). *WAIS-R manual: Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: Psychological Corporation.

Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N. R., Chui, H., … Morris, J. C. (2009). The Alzheimer's Disease Centers' Uniform Data Set (UDS): The neuropsychological test battery. *Alzheimer Disease and Associated Disorders*, *23*, 91–101.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, *38*, 330–336.

Welch, L. W., Doineau, D., Johnson, S., & King, D. (1996). Educational and gender normative data for the Boston Naming Test in a group of older adults. *Brain and Language*.

Welsh, K. A., Butters, N., Hughes, J. P., Mohs, R. C., & Heyman, A. (1992). Detection and staging of dementia in Alzheimer's disease: Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Archives of Neurology*, *49*, 448–452.

Whitfield, K. E., Fillenbaum, G. G., Pieper, C., Albert, M. S., Berkman, L. F., Blazer, D. G., … Seeman, T. (2000). The Effect of race and health-related factors on naming and memory: The MacArthur Studies of Successful Aging. *Journal of Aging and Health*, *12*, 69–89.

Wilkins, J. W., Hamby, S. L., & Thompson, K. L. (1996). Difficulties with Boston Naming norms in individuals with below average WAIS-R vocabulary. *Archives of Clinical Neuropsychology*, *11*, 464–464.

Williams, B. W., Mack, W., & Henderson, V. W. (1989). Boston Naming Test in Alzheimer's disease. *Neuropsychologia*, *27*, 1073–1079.

Williams, S. E., & Canter, G. J. (1982). The influence of situational context on naming performance in aphasic syndromes. *Brain and Language*, *17*, 92–106.

Williams, V. G., Bruce, J. M., Westervelt, H. J., Davis, J. D., Grace, J., Malloy, P. F., & Tremont, G. (2007). Boston naming performance distinguishes between Lewy body and Alzheimer's dementias. *Archives of Clinical Neuropsychology*, *22*, 925–931.

World Bank Development Data Group (Eds.). (2012). *World development indicators 2012*. Washington, DC: World Bank-free PDF.

Worrall, L. E., Yiu, E., Hickson, L. M., & Barnett, H. M. (1995). Normative data for the Boston Naming Test for Australian elderly. *Aphasiology*, *9*, 541–551.

Yeates, K. O., Taylor, H. G., Wade, S. L., Drotar, D., Stancin, T., & Minich, N. (2002). A prospective study of short-and long-term neuropsychological outcomes after traumatic brain injury in children. *Neuropsychology*, *16*, 514.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of psychiatric research*, *17*, 37–49.

Zec, R., Burkett, N., Markwell, S., & Larsen, D. (2007a). A cross-sectional study of the effects of age, education, and gender on the Boston Naming Test. *The Clinical Neuropsychologist*, *21*, 587–616.

Zec, R., Burkett, N., Markwell, S., & Larsen, D. (2007b). Normative data stratified for age, education, and gender on the Boston Naming Test. *The Clinical Neuropsychologist*, *21*, 617–637.

Zec, R., Markwell, S., Burkett, N., & Larsen, D. (2005). A longitudinal study of confrontation naming in the "normal" elderly. *Journal of the International Neuropsychological Society*, *11*, 716–726.

Zhou, X., Obuchowski, N. A., & McClish, D. K. (2002). *Statistical methods in diagnostic medicine*. New York: Wiley Interscience.

Appendix A

Demographic Questionnaire for Control Participants

DEM - Older Adult English Demographic Questionnaire

GENERAL INFORMATION

Full name:			
Date of Birth:	YYYY	MM	DD
Test Date:	YYYY	MM	DD
What is your gender?	1. Male 2. Female		
How would you describe your race?	1. Black 2. Colored 3. White 4. Asian 5. Other(specify): 6. Refuse to answer		
What language(s) do you speak?	1. English 2. Afrikaans 3. Xhosa 4. Zulu 5. Other (specify):		
Which do you use most often?	1. English 2. Afrikaans 3. Xhosa 4. Zulu 5. Other (specify):		

	Person	Home	Work	Cell
Contact numbers:	Self			
	Cohabitant			
Residential Address:				

RESIDENTIAL INFORMATION

How long have you lived at your current address?	
How would you describe your dwelling?	1. Shack 2. Wendy house or backyard dwelling 3. Tent or traditional dwelling 4. Flat / apartment 5. Town house / semi-detached house 6. Freestanding brick house 7. Other (specify):
Which of these items do you have in your home? (mark as many as necessary)	A. Tap water B. Flush toilet inside home C. Electricity D. Telephone (landline) E. Television F. Computer G. Car

How many people, other than yourself, sleep in the same room with you at night when you are at home?	1. One 2. Two 3. Three 4. Four 5. Five 6. More than five 7. None

EDUCATION

Name and area of Primary School attended:	School: Suburb / area:
Name and area of Secondary School attended:	School: Suburb / area:
What is the highest grade of school education you completed?	
Name of Tertiary Institution (if attended):	
How many years of tertiary education have you completed (and passed completely?)	

MEDICAL HISTORY

Do you have any problems with your sight, hearing or with co-ordination?	1. No 2. Yes
If YES, please provide some details:	

Have you ever been admitted to hospital?	1. No 2. Yes (If YES, please answer the following)
Why were you hospitalized?	
How old were you?	
How long did you stay in hospital?	

Have you ever had a head injury?	1. No 2. Yes (If YES, please answer the following)
How did the injury occur?	
Did you lose consciousness?	
How long were you unconscious?	
How old were you?	

Have you ever had a fit / seizure?	1. No 2. Yes (If YES, please answer the following)
How old were you?	
What caused it?	

Has it happened more than once?	
Do you take medication for it?	

Have you ever had a serious illness?	1. No 2. Yes (If YES, please answer the following)	
Name of illness/es		Age

Have you ever had to take medication for over two weeks? (do not include medication for common conditions such as colds, flu, gastro enteritis)	1. No 2. Yes (If YES, please answer the following)
What was the reason for the medication?	
What was the name and dosage of the medication?	
Are you currently taking any medication?	
What is the reason for the medication?	
What is the name and dosage of the medication?	

PSYCHIATRIC HISTORY

Have you ever sought counseling (at school, church or elsewhere) for emotional or other difficulties?	1. No 2. Yes (If YES, please answer the following)
How old were you?	
Who did you receive help from?	
For how long did you consult the person / agency?	
Did the treatment help your condition?	

PSYCHOMETRIC HISTORY

Have you had a psychometric evaluation (for example, an "IQ" test) in the last 12 months?	1. No 2. Yes (If YES, please answer the following)
What was the purpose of the test?	
Who tested you?	

SCHOLASTIC HISTORY

In comparison with your peer group, have you ever experienced severe difficulties in coping with your school work?	1. No 2. Yes (If YES, please answer the following)
If YES, please provide some details?	

SOCIO-ECONOMIC INFORMATION

What is the total monthly income of your household?	1. Less than R499 2. R500 – R999 3. R1000 – R1999 4. R2000 – R4999 5. R5000 – R9999 6. More than R10 000
Are you retired?	1. Yes 2. No
What is/was your occupation?	

Thank You

Appendix B

BNT-SA-SF Scoring Sheet

BNT Short Form

INSTRUCTIONS: Ndizakubonisa imifanekiso ethile. Ndicela undixelele into oyibonayo kumfanekiso ngamnye.
I am going to show you some pictures. I would like you to name the thing that you see in each picture.
Ek gaan nou vir jou 'n paar prentjies wys. Ek wil hê dat jy die ding wat jy in elke prentjie sien, noem.

umthi *tree* **boom**		√ or *X*	verbatim response (if *X*):	error code
1	**Uncued response**	√ or *X*		
	Stimulus cue: into ekhulela phandle *something that grows outdoors* iets wat buite groei	√ or *X*		
	Phonemic cue: umth… *tr…*b…	√ or *X*		
	Multiple choice	√ or *X*	igqabi intyatyambo isityalo **umthi**	
			leaf flower plant ***tree***	
			blaar blom plant **boom**	

ikama *comb* **kam**		√ or *X*	verbatim response (if *X*):	error code
2	**Uncued response**	√ or *X*		
	Stimulus cue: isetyenziselwa ukulungisa iinwele *used for fixing hair* word gebruik om hare mee netjies te maak	√ or *X*		
	Phonemic cue: ika… *co…* ka…	√ or *X*		
	Multiple choice	√ or *X*	iinwele iharika **ikama** ibrash	
			hair rake ***comb*** *brush*	
			hare hark **kam** borsel	

itoothbrush _toothbrush_ tandeborsel	√ or _X_	verbatim response (if _X_):	error code
3 **Uncued response**	√ or _X_		
Stimulus cue: isetyenziswa emlonyeni _used in the mouth_ word in die mond gebruik	√ or _X_		
Phonemic cue: itoo… _too…_ ta…	√ or _X_		
Multiple choice	√ or _X_	**itoothbrush** amazinyo umsonto intlama yamazinyo	
		toothbrush _teeth_ _floss_ _toothpaste_	
		tandeborsel tande vlos tandepasta	

ihengari _hanger_ hanger	√ or _X_	verbatim response (if _X_):	error code
Uncued response	√ or _X_		
Stimulus cue: ifunyanwa ekhabathini _found in a cupboard_ word in 'n kas gevind	√ or _X_		
4 **Phonemic cue:** ihe… _ha…_ ha…	√ or _X_		
Multiple choice	√ or _X_	indawo yokubamba **ihengari** ikhabathi iimpahla	
		handle **_hanger_** _cupboard_ _clothes_	
		handvatsel **hanger** kas klere	

5	**ibhanki** *bench* **bank**	√ or **X**	verbatim response (if **X**):				error code
	Uncued response	√ or *X*					
	Stimulus cue: isetyenziselwa ukuhlala word gebruik om op te sit	√ or *X*					
	Phonemic cue: ibha… *be…* ba…	√ or *X*					
	Multiple choice	√ or *X*	isitulo	izitepsi	**ibhanki**	ipaki	
			chair	*steps*	***bench***	*park*	
			stoel	trappe	**bank**	park	

6	**inkumba-ndonda** *snail* **slak**	√ or **X**	verbatim response (if **X**):				error code
	Uncued response	√ or *X*					
	Stimulus cue: isilwanyana *an animal* 'n dier sl…	√ or *X*					
	Phonemic cue: ink… *sn…* *sl…*	√ or *X*					
	Multiple choice	√ or *X*	inyoka	iqokobhe **inkumba-ndonda**	ufodo		
			snake	*shell*	***snail***	*turtle*	
			slang	skulp	**slak**	skilpad	

	idathi *dart* veerpyltjie	√ or *X*	verbatim response (if *X*):	error code
7	**Uncued response**	√ or *X*		
	Stimulus cue: uyayijula *you throw it* 'n mens gooi dit	√ or *X*		
	Phonemic cue: ida… *da…* vee…	√ or *X*		
	Multiple choice	√ or *X*	utolo **idathi** unotaka umkhonto *arrow* **dart** *pin* *spear* pyl **veerpyltjie** speld spies	

	umkhombe / irhino *rhinoceros* renoster	√ or *X*	verbatim response (if *X*):	error code
8	**Uncued response**	√ or X		
	Stimulus cue: isilwanyana *an animal* 'n dier	√ or X		
	Phonemic cue: umkho… / irhi… *rhi…* re…	√ or X		
	Multiple choice	√ or X	imvubu **umkhombe / irhino** inyathi indlovu *hippopotamus* **rhinoceros** *buffalo* *elephant* seekoei **renoster** buffel olifant	

iidomino *dominoes* domino's	√ or X	verbatim response (if X):	error code
Uncued response	√ or X		
Stimulus cue: umdlalo *a game* 'n speletjie	√ or X		
Phonemic cue: iido… *do…* do…	√ or X		
Multiple choice	√ or X	idayisi ichess umdlalo **iidomino** *dice chess game **dominoes*** dobbelstene skaak speletjie **domino's**	

The left cell of the first table row spans the number **9**.

izitepsi ezihambayo *escalator* roltrap	√ or X	verbatim response (if X):	error code
Uncued response	√ or X		
Stimulus cue: uyenyuka kuyo *you go up on it* 'n mens gaan daarmee op	√ or X		
Phonemic cue: izit.. *es…* ro…	√ or X		
Multiple choice	√ or X	ilift isingxobo sekatala **izitepsi ezihambayo** izitepsi *elevator guitar case **escalator** stairs* hysbak kitaartas **roltrap** trap	

The left cell of the second table row spans the number **10**.

istethaskophu *stethoscope* stetoskoop	√ or X	verbatim response (if X):	error code
	Uncued response	√ or X		
	Stimulus cue: isetyenziswa ngooqgirha nabongikazi *used by doctors and nurses* word deur dokters en verpleegsters gebruik	√ or X		
11	**Phonemic cue:** iste... *ste...* ste...	√ or X		
	Multiple choice	√ or X	itheleskophu izimameli isishushubezi-ntliziyo **istethaskophu**	
			*telescope earphones heartbeater **stethoscope***	
			teleskoop oorfone hartklopper **stetoskoop**	

ifanele *funnel* tregter	√ or X	verbatim response (if X):	error code
	Uncued response	√ or X		
	Stimulus cue: isetyenziselwa ukugalela *used for pouring* word gebruik om te skink	√ or X		
12	**Phonemic cue:** ifa... *fu...* tre...	√ or X		
	Multiple choice	√ or X	isihluzo igilasi **ifanele** umbhobho	
			*filter glass **funnel** siphon*	
			filter glas **tregter** sifon	

ikhamphas *compass* passer	√ or X	verbatim response (if X):	error code
Uncued response	√ or X		
Stimulus cue: ukuzoba *for drawing* om mee te teken	√ or X		
Phonemic cue: ikha… *co…* pa…	√ or X		
Multiple choice	√ or X	iprotrakta izahluli **ikhampas** izangqa *protractor dividers **compass** circles* gradeboog verdeelpasser **passer** sirkels	

(Row number 13 appears to the left of this table)

isfinks *Sphinx* Sfinks	√ or X	verbatim response (if X):	error code
Uncued response	√ or X		
Stimulus cue: ifunyanwa eYiphutha *it's found in Egypt* dit kom in Egipte voor	√ or X		
Phonemic cue: isfi… *sph…* sf…	√ or X		
Multiple choice	√ or X	**isfinks** istetyhu ufaro iphiramidi ***Sphinx** statue pharaoh pyramid* **sfinks** standbeeld farao piramide	

(Row number 14 appears to the left of this table)

	iprotrektha *protractor* **gradeboog**	√ or X	verbatim response (if X):	error code
15	**Uncued response**	√ or X		
	Stimulus cue: ilinganisa ii-engile *measures angles* meet hoeke	√ or X		
	Phonemic cue: ipro… *pro…* gra…	√ or X		
	Multiple choice	√ or X	ispidomitha irula ikhampas **iprotrektha** *speedometer ruler compass **protractor*** snelheidsmeter liniaal passer **gradeboog**	

Appendix C

Letter of Introduction to Social Housing Company

Department of Psychology
P.D. Hahn Building
University of Cape Town
Rondebosch
7701

1 October, 2012

██████████████
██████████

To whom it may concern

University of Cape Town Research

I am a researcher from the Department of Psychology at the University of Cape Town. As you will be aware, changes in memory and the ability to name people and things that may be familiar are a normal part of aging. However, there are a number of tests that we can use to differentiate normal aging from that associated with real, underlying problems. We are conducting research into how tests from overseas function in the South African population, in an effort to design tests that are suitable for use with our diverse population.

We are currently recruiting healthy older adults in the Cape Town region to participate in this study and have identified ██████████ facilities in the Southern Suburbs, such as █ ████████ as a possible location. Should you agree, we would come to the facility on a day and time that suits you and hold a general information session for all residents that are interested in hearing more about our research. Should they wish to take part, and fit the criteria, we would then set up individual appointments with them to administer the four short "tests". These "tests" are not tests like you do at school, nor are they intelligence tests. The participant will be asked to answer a few questions and do some activities like naming things or describing things.

This would be a great opportunity to not only learn more about research that is relevant to the lives of older adults, but also to be screened for possible serious underlying conditions that have not been detected. Should the individual's performance on the tests indicate a possible problem, they will be referred to the Memory Clinic at Groote Schuur hospital for an appointment with medical professionals, should they wish to follow this up. In addition, all participants will be entered into a raffle to win one of two shopping vouchers to a supermarket of their choice.

Should you require more information, please do not hesitate to contact me.

Kind regards,
Lauren Baerecke
M.A. Candidate
Department of Psychology
University of Cape Town

Appendix D

Flyer Advertising Study

Dear Resident,
You are invited to take part in a study with researchers from the University of Cape Town. As you may be aware, as people get older they might start to see changes in their memory and forget the names of people and things that may be familiar to them. We are doing research around these issues.

<u>Who can take part?</u>
Anybody can take part in this study as long as you are an adult over 50 years of age, speak English and/or Afrikaans and are generally healthy.

<u>What does it involve?</u>
A researcher will set up an appointment with you at a time and place that suits you, even in your own home. They will conduct a 45-minute interview with you during which you will be asked to answer some questions and do some activities like naming things.

If you would like to participate in this research or would just like to know more about the research we are doing, please attend the general information session which will be held at ▮▮▮▮ on
25 September 2012
at 10am

If you cannot attend the meeting but would still like to participate, please phone Lauren on ▮▮▮▮▮

Appendix E

English Consent Form for Control Participants

<u>**ENGLISH ADULT CONSENT AND INFORMATION LEAFLET**</u>

Dear Volunteer

Why is this study being done?
A research study is a way to learn more about something. A team of researchers from the University of Cape Town are trying to learn more about the ways in which people's brains develop. Many of the tests (called neuropsychological tests) that we use come from other countries, like the United States. To be able to use these tests in South Africa, for people who speak English, Afrikaans or Xhosa, we need to investigate how ordinary South Africans perform on these tests.

Who can take part in the study?
Adults who are older than 50 years may take part in this study. If you speak English or Afrikaans as your home language, you may take part. If you have had serious problems with your health or schooling, or have been diagnosed with dementia or any other mental illness, you will not be able to take part in this study. However, you may ask the researcher about taking part in some of our other studies.

What will happen to you if you agree to take part in this study?
If you agree to take part in the study, you will be asked to sign this form and fill out a demographic questionnaire which asks you some questions about yourself, your education, and your medical history. A researcher will then make an appointment to come and meet with you at a time and place that suits you. At the visit, the researcher will explain everything to you and answer any questions or concerns you may have. Then the researcher will administer four simple and straightforward assessments or "tests" to you, by yourself. You will be asked to answer a few questions and do some activities like naming things or describing things to us. You do not have to study for the "test" and they are not testing your intelligence or how much you know. In fact, you are not expected to get everything correct. All you will be asked to do is to try your best. It will take about half an hour to complete. At any time during the procedure, you may ask the researcher to stop if you decide you no longer want to participate.

What will happen to the information you give us?
This study is confidential and if you agree to take part, the researcher will use a code for the information about you. This means that all the information that you give us will be kept private and your name will never be used.

Do you have to pay to take part?
You don't have to pay anything to take part in this study. If you fill in all the forms and complete the tests, you will be entered into a raffle to win a prize. This is our way of thanking you for taking part in the study.

Do you have to take part in the study?
You do not have to take part in the study. It is up to you to decide whether you want to take part or not. If you want to take part, then please write your name on this form, and sign it. If

you sign the papers now and then decide to change your mind later, all you have to do is to tell us that you don't want to take part anymore. There will be no consequences should you change your mind or decide not to take part.

What if you have any questions?

If you have any questions about this study, you may ask the researcher about them during one of the visits, or later on. You can phone Lauren Baerecke between 08h30 and 16h30 on weekdays on ▮▮▮▮▮▮ if you have any more questions.

If you agree to take part in this study and you understand what has explained to you, please write and sign your name below:

Name: _____

Signature: _____

Date: _____ Place: _____

The researcher must sign his/her name below to confirm that he/she has explained the study to you and answered the questions you have about it:

Researcher's name: _____

Researcher's signature: _____

Date: _____ Place: _____

Appendix F

Figures Presenting Key Values For Calculating Diagnostic Efficiency Statistics

Diagnosis

		Present	Absent
	Positive	a 37	b 5
Test			
	Negative	c 9	d 46

Figure F1. True positive, false positive, true negative, and false negative values for the AD verses controls comparison.

Diagnosis

		Present	Absent
	Positive	a 21	b 23
Test			
	Negative	c 2	d 28

Figure F2. True positive, false positive, true negative, and false negative values for the OD verses controls comparison.

Diagnosis

		Present	Absent
	Positive	a 27	b 4
Test			
	Negative	c 19	d 19

Figure F3. True positive, false positive, true negative, and false negative values for the OD verses AD comparison.

Appendix G

Table Presenting Revised Order of BNT-SA-SF Items

Table G.
Revised Order of BNT-SA-SF Items Based on Item Difficulty for the Control Group (*N = 51*)

Item	Recommended position	Original position	Difficulty index (%)
Tree	Item 1	Item 1	100
Comb	Item 2	Item 2	100
Toothbrush	Item 3	Item 3	100
Hanger	Item 4	Item 4	100
Bench	Item 5	Item 5	100
Dart	Item 6	Item 7	100
Snail	Item 7	Item 6	98
Funnel	Item 8	Item 12	98
Rhinoceros	Item 9	Item 8	96.1
Escalator	Item 10	Item 10	96.1
Stethoscope	Item 11	Item 11	94.1
Dominoes	Item 12	Item 9	92.2
Sphinx	Item 13	Item 14	52.9
Compass	Item 14	Item 13	33.3
Protractor	Item 15	Item 15	25.5

Note. Difficulty index is the proportion of individuals who produced the correct response spontaneously or following a semantic cue.