Reliability and Validity of a South African Adaptation of the Boston Naming Test

Lauren Baerecke
ACSENT Laboratory
Department of Psychology
University of Cape Town

Supervisor: Dr Kevin Thomas
Co-supervisor: Helen Ferrett

Word Count:

Abstract: 235

Main Body (excluding tables and graphs): 9895

Abstract

This research examined the psychometric properties of a modified short form of the Boston Naming Test (BNT), as part of a larger project aiming to modify, translate, and norm commonly used neuropsychological tests in South Africa. This modified BNT includes 15 items, drawn from the pool of items on the original test, judged more culturally appropriate for the South African population than those on other commonly used short forms. Study 1 demonstrated the cultural bias inherent in the original 60-item BNT, by comparing the performance of normal healthy South African university students with a high level of education to North American normative data. The performance of this sample on the modified short form, including reliability measures, was also investigated, and indicated the chosen test items were not problematic in this population group. Study 2 focused on validity. Preliminary cross-cultural validity of the modified short form was examined by comparing the performance of two groups of healthy community-dwelling older adults. Results showed that Xhosa-speaking individuals of low socioeconomic status (SES) performed significantly more poorly than English-speaking individuals of high-SES did. Further, the diagnostic ability of the modified test was compared to a widely used short form, and found that, in a South African memory clinic population, patients with Alzheimer's disease and vascular dementia performed similarly on both versions.

Keywords: Boston Naming Test; Boston Naming Test short form; cultural bias; cross-cultural neuropsychology; South Africa; normative data

Word-finding difficulties are a common complaint amongst older adults. It is important, however, for clinicians to distinguish between the kinds of mild word-finding difficulties found in healthy elderly people, and the more serious word-finding difficulties that reflect underlying cognitive or neurological impairments found in age-related disease processes, such as dementia of the Alzheimer's (AD) or vascular type (VaD). 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 (Lezak, Howieson, & Loring, 2004; Williams, Mack, & Henderson, 1989).

Numerous studies have examined the issues that arise when using the test outside of North America. The identification of cultural bias has led to the modification of the BNT for use with different languages and cultures. Despite the clinic popularity of the BNT and its various short forms, 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 a widely used test of confrontation naming ability. Confrontation naming is defined as "the ability to pull out the correct word at will" (Lezak et al., 2004, p. 511). An experimental version of the BNT was first published in 1978 and two revisions followed, the most recent featuring the 60 items that are in common use today (Kaplan et al., 1983, 2001).

In its current form, the BNT consists of 60 black-and-white line drawings that are presented in ascending order of difficulty, from common items such as *bed* and *toothbrush* to less familiar items such as *protractor* and *palette*. Examinees are given 20 seconds to produce a correct spontaneous response, after which a semantic cue is offered. Failing the production of a 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*'.

The most recent revision of the instrument also features the addition of a multiple-choice section. In this section, after completing the standard presentation as described above, the examiner returns to each item the examinee has failed. The examinee is asked to select the correct choice, i.e. the word that best describes the pictured object, from four multiple-choice options read aloud by the examiner.

Development of BNT Short Forms

A number of 15- and 30-item BNT versions have been developed (Fastenau, Denburg, & Mauer, 1998; Graves, Bezeau, Fogarty, & Blair, 2004; Lansing, Ivnik, Cullum, & Randolph, 1999). Most notably, Mack, Freed, Williams, and Henderson (1992) developed four equivalent 15-item versions by dividing the 60 items of the original test into four 15-item groups, each reflecting the full range of content of the full test. The authors' fourth version, the Mack SF-4, is the short form included with the BNT kit. Another popular short form is the Williams 30-item test, which was empirically derived by choosing items that best discriminated between patients with AD and healthy controls (Williams et al., 1989).

These shorter tests are useful for several reasons. Time constraints in clinical and research settings mean that administering the full 60-item BNT is impractical in some situations. In other words, short forms aid in the rapid screening of patients (Mack et al., 1992). 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). Lastly, 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).

Psychometric Properties of BNT Short Forms

Reports on the psychometric properties of the numerous short forms vary. Psychometric studies of 30-item forms have reported good reliability and high correlations with the full test (Graves et al., 2004; Saxton et al., 2000). However, reported reliability coefficients and correlations with the BNT for 15-item versions are lower than for the 30-item versions (Strauss, Sherman, & Spreen, 2006). This discrepancy might arise because many of the 15-item forms are developed to retain content validity, i.e., to preserve the full content range with regards to item difficulty. Those short forms that display better psychometric properties have been developed for specific purposes, such as identifying AD (Graves et al., 2004).

Despite the value of these abbreviated versions, there is little normative data available for them. Without investigation of the psychometric properties and the provision of normative data, the clinical and research utility of these tests is limited (Jefferson et al., 2007).

Clinical Findings: Distinguishing Patients with AD

The diagnostic information provided by the BNT, whether in full or short form, is used in the assessment of naming deficits in patients of all ages with cognitive or neurological problems including aphasia, acquired brain injury, and degenerative disorders (Diehl et al., 2005; Larrain & Cimino, 1998).

Although there is controversy regarding the mechanism underlying naming deficits in AD , numerous studies have shown the effectiveness of the 60-item BNT in detecting the naming deficits present in this dementia, and thus distinguishing AD from normal aging (Mack et al., 1992; Williams et al., 1989). For instance, Lansing et al. (1999) showed a significant difference in total BNT score between 325 AD patients and 719 elderly normal control participants.

Additionally, AD patients tend to perform more poorly than VaD patients on the BNT. Barr, Benedict, Tune, and Brandt (1992) identified the BNT as one of the most discriminative tests in a neuropsychological battery identifying patients with either of these two forms of dementia. It has been found that, as well as producing more general errors, AD patients also differ from VaD patients in the quality of semantic errors produced (Lukatela, Malloy, Jenkins, & Cohen, 1998). Particularly, AD patients tend to produce superordinate category responses, such as 'animal' instead of 'rhinoceros'. There are conflicting findings, however. For instance, Mathias and Burke (2009) point to the difficulty in differentiating between AD and VaD using the BNT, particularly as more recent research has disputed the traditional understanding that these two dementias result from different etiologies. Using a meta-analysis, these authors found that the BNT did not successfully discriminate between AD and VaD.

Of crucial importance in establishing the validity of a BNT short-form is its ability to classify impaired subjects in a clinical study correctly. Various BNT short forms have been found to be a high rate of agreement with the long form in correctly classifying AD/VaD patients as abnormal (Graves et al., 2004; Jefferson et al., 2007). Mack et al. (1992) found that the four 15-item versions they had developed successfully differentiated a sample of 26 AD patients from 26 healthy controls. However, using the same sample as described above, Lansing et al. (1999) found that the Mack short forms did not discriminate between controls and AD patients as well as the 60-item version did. There is little available evidence of the effectiveness of BNT short forms in discriminating between AD and VaD. Therefore, although short forms are useful in a number of situations, the evidence as to their clinical validity varies.

In summary, the BNT has been widely used as a diagnostic tool in distinguishing patients with AD. Due to the widespread use of the BNT in detecting even mild naming impairments, many authors have identified the importance of investigating non-neurological factors that may impact on 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 impact on scores. The literature identifies four such factors: age, education, sex, and culture.

Age and BNT performance. Age has frequently been identified as one of the primary demographic factors that moderates BNT performance in healthy individuals (Lucas et al., 2005; Saxton et al., 2000; Zec, Burkett, Markwell, & Larsen, 2007a). These studies have found that BNT performance declines with increasing age. In the largest such study (*N*= 1172 healthy individuals), Zec et al. (2007a) showed significantly poorer mean scores and greater variability in scores in successively older age groups between 50 and 99 years. These findings are reasonably consistent with those from other cross-sectional studies (e.g., Cruice, Worrall, & Hickson, 2000). Some studies have, however, pointed to evidence that larger, significant age differences only appear after the age of 70 (Tombaugh & Hubley, 1997) or 80 years (Kent & Luszcz, 2002).

Education and BNT performance. Many studies report a positive association between lower scores and lower levels of education (Weintraub et al., 2009; Worrall, Yiu, Hickson, & Barnett, 1995; Zec, Burkett, Markwell, & Larsen, 2007b). In a review of the relationship between education and BNT performance, Hawkins and Bender (2002) found that education effects appear to be strongest at lower education levels (< 12 years). In other words, education effects are typically not found where participants have 12 or more years of education, but the test does distinguish between higher and lower education levels.

Sex and BNT performance. Research findings on sex differences in BNT performance have been less consistent than those reporting age or education effects. Although some early studies reported sex differences in favour of men in older samples (Tombaugh & Hubley, 1997), more recent studies have not found such differences (Saxton et al., 2000; Zec et al., 2007b).

Similar findings regarding the effects of age, education, and sex have been replicated with different short forms (Graves et al., 2004; Kent & Luszcz, 2002; Saxton et al., 2000) and across numerous countries, within different cultures and languages, including Korea (Kim & Na, 1999), Sweden (Tallberg, 2005), and Australia (Worrall et al., 1995).

Culture and BNT performance. A fourth source of potential influence on BNT scores is the cultural relevance of items to the test-takers. The BNT, as with many other popular neuropsychological tests, was developed for the assessment of English-speaking, North American individuals, and reflects the context in which it was developed. Performance on the BNT by individuals from non-North American cultures has been found to differ significantly from the performance of North American participants, and many authors have pointed to the potential for misdiagnosis of naming deficits in such contexts (Barker-Collo, 2001, 2007; Cruice et al., 2000; Kim & Na, 1999).

A cross-cultural effect on BNT performance has been found within different ethnic groups within the United States. For instance, Pedraza et al. (2009) found significant differences between white Americans and African-Americans on the BNT. Similar findings have emerged from studies on Spanish-speaking samples in the US (de la Plata et al., 2009). It appears that, with regards to different ethnic groups, those who are most acculturated to the dominant US culture score significantly higher than those less familiar with that culture (Strauss et al., 2006).

The same applies to different populations. Canadians have been found to perform as well as Americans (Tombaugh & Hubley, 1997) but in other English-speaking countries, such as New Zealand and Australia, BNT performance of normal healthy individuals was found to be significantly worse than that of comparable North American individuals (Barker-Collo, 2001; Worrall et al., 1995). Thus, caution is warranted when administering the BNT to individuals outside of North America (even if those individuals have English as a home language) because of the potential for cultural bias if individuals are not familiar with the culture in which the test was developed.

The literature not only identifies between-group differences in mean BNT scores, but also finds differences in the pattern of errors made by examinees with different cultural backgrounds. Particular items have been identified as being especially culturally loaded. For instance, in both younger and older adult 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; Worrall et al., 1995).

Such findings have led to suggestions that the test should be appropriately modified to replace such problematic items when it is used outside of North America (Cruice et al., 2000). Following this suggestion, researchers and clinicians across the world have developed a number of culturally modified versions of the test.

Adaptation of the BNT for Use with Different Cultures and Languages

Much of the literature on the use of the BNT outside of North America revolves around adaptations of the test for use with other cultures and languages. The enduring popularity and utility of the BNT is suggested by the fact that it has been translated and adapted for use within different cultures and languages, including Greek (Patricacou, Psallida, Pring, & Dipper, 2007), Swedish (Tallberg, 2005), and Korean (Kim & Na, 1999).

Different approaches have been adopted in the translation and cultural adaptation of the BNT. When modifying the test for the Swedish population, Tallberg (2005) changed the sequence of items but did not replace any items (i.e., did not take into account potential cultural bias of particular items). In their modification of the BNT for use with the Greek population, Patricacou et al. (2007) used a pilot test that identified four problematic items (*pretzel, doorknocker, stethoscope* and *scroll*); these items were replaced with traditionally Greek items from the same semantic categories. In their development of a Korean version (the K-BNT), Kim and Na (1999) included only 10 items from the original BNT due to extensive cultural differences.

In summary, there is consistent evidence that the BNT needs to be modified if it is to be used outside of North America. Although the predominant trend has been to replace items that are identified as culturally biased with more appropriate items (Kim & Na, 1999; Patricacou et al., 2007; Worrall et al., 1995), some modifications have attempted to use items from the original test. 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 reliable and valid neuropsychological measures to be used outside of the country in which they were developed.

Current Issues in Cross-cultural Neuropsychology in South Africa

Clinicians working in various settings in developing countries, including South Africa, are increasingly using standard neuropsychological tests developed in the global north in their assessments of patients with cognitive or neurological impairment. However, without locally-appropriate normative data or new, culture-fair tests, clinicians working in these developing-world settings may struggle in the accurate assessment of their patients (Nell, 2000; Shuttleworth-Edwards et al., 2004).

Researchers have approached the issue of cross-cultural neuropsychological testing in at least two ways. They either use tests that are cross-culturally fair, which may mean having to create new tests, or they create locally-appropriate normative data, taking into consideration demographic and cultural factors (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, culturally fair tests (Shuttleworth-Jordan, 1996). In South Africa, researchers conducting norming studies must acknowledge that performance on neuropsychological tests is moderated by numerous demographic variables, including SES and education (Nell, 2000). Unfortunately, the number of published research studies presenting such data on the BNT for the South African population is woefully inadequate.

Summary and Rationale for the Present Study

Neuropsychological tests developed and normed in Western and/or North American countries can be misleading when used in developing countries, such as South Africa (Nell, 2000). There is therefore a desperate need amongst South African clinicians for neuropsychological tests appropriate for assessment in the local context, as well as relevant normative data, which is essential to ensure the diagnostic utility of such tests (Shuttleworth-Jordan, 1996). The scope for research in this field is therefore wide.

The BNT, a popular and widely-used confrontation naming test, is already used in clinical settings and research studies in South Africa, despite the fact that it may be inappropriate because of the established cross-cultural issues. With the development of neuropsychology in South African as a profession, where increasing numbers of neuropsychologists are being trained and will be working in this field, the need for valid and reliable tests for use with the South African population is growing. In addition, with increasing amounts of overseas grant money being invested into South African-based research into HIV and other conditions, the need for reliable instruments that are also well known to international audiences is great.

Although some studies conducted outside of North America have developed and validated the use of modified BNTs for use within that specific culture or language, until recently no published research on the use of BNT in South Africa was available for clinicians (see Mosdell, Balchin, & Ameen, 2010, for a recent paper on the development of a BNT version suitable for inpatient use in South Africa). Thus, researchers from the University of Cape Town and the University of Stellenbosch have 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 has led to the development a modified short form, the BNT-SA-SF, which includes 15 items judged by a forum of practicing neuropsychologists as being more culturally appropriate for the South African population than those on the most popular 15-item short form (Mack SF-4; Mack et al., 1992) . The current study was a preliminary investigation into the psychometric properties of this modified short form.

Study 1

Study 1 investigated the performance of a sample of undergraduate students who were fluent in English, emerged from high SES backgrounds, and had experienced a high quality of education. The specific questions to be addressed were: (1) How does this sample perform compared to the North American normative sample on the BNT, (2) how does the modified test (BNT-SA-SF) fare in terms of test-retest reliability, internal consistency, and its correlation with the BNT when evaluated in the current South African sample.

Methods

Design and Setting. A quantitative, single-sample design with repeated measures was used in accordance with nonrandomized selection criteria. There were two administrations of the BNT with a 2-week interval period to determine test-retest reliability.

Data were collected from UCT students. Data collection took place in the ACSENT laboratory, in the UCT Department of Psychology.

Participants. This study included 30 UCT undergraduate students between the ages of 18 and 25 years. They participated in exchange for course credit, via the Student Research Participation Program (SRPP). The eligibility criteria for participation in the study process were applied strictly to ensure the desired demographic composition of the sample and to ensure results were not confounded by extraneous variables.

Those eligibility criteria emerged from the fact that I hypothesised (based on previous literature in the field) that healthy English-speaking young adults with a high level and quality of education, and who emerged from high SES backgrounds, were most likely to achieve the highest scores on the 60-item BNT and thus perform most similarly to North American normative samples. Another contributing factor to the choice of this demographic group was the fact that such individuals would also likely have been exposed to American culture via their education, the media and Internet.

Specifically, the eligibility criteria were these: Participants were required to speak English as their first language and to be a UCT undergraduate student. They were also required to have matriculated from a high school that fell within the fourth or fifth quintiles of economic status according to 2001 census data; that is, only participants who had matriculated from a school in the highest or second-highest economic brackets were included in the sample. The median annual income of each participant's suburb of residence, again determined according to 2001 census data, was used to determine if he/she was of high SES. A cut-off of R150 000 was used so that only participants whose suburbs of residence had a mean annual income equal to or higher than that number were included in the sample.

To ensure participants were healthy, an extensive self-report medical history was obtained from each. The following criteria warranted exclusion from the study: current use of psychotropic medication, and/or a history of psychiatric diagnosis, pre-natal or birth complication, a head injury that resulted in a loss of consciousness for more than 5 minutes, seizure disorders, substance abuse disorders, a medical illness resulting in loss of cognitive functioning, and language, speech or behavioural disorders. Additionally, having received psychometric testing that involved intelligence testing in the past 12 months warranted exclusion from this study. These criteria were used because they have been shown to influence performance on neuropsychological tests, including the BNT (Mitrushina, Boone, Razani, & D'Elia, 2005).

The final sample included 15 males and 15 females. All were white, and all met the eligibility criteria. Table 1 shows descriptive statistics for the key demographic variables of the sample.

Table 1
Demographic Characteristics of the Current Sample

Variable	Range	*M* (*SD*)
Age (years)	18.33-24. 50	21.08 (1.89)
Education	12-17	13.00 (1.29)
SES	150963.00-383522.00	274981.83 (75920.61)

Note. The variable *Age* was calculated from participant's age in months at test time 1. That number was divided by 12 to achieve a decimal score. The variable *Education* reflects the number of years successfully completed. The variable *SES* was estimated using 2001 census data for the median annual income level in Rands for each participant's suburb of residence.

Measures. All participants completed a consent document (see Appendices A) and a demographic questionnaire (see Appendix B).

The standard 60-item BNT was administered to all participants according to the standardized instructions that appear in the test manual, with this exception: All 60 items were presented (i.e., administration began at item 1 and continued through item 60; starting point and discontinuation rules were not applied). This procedure was followed to ensure that performance on all 60 items could be examined

The BNT-SA-SF comprises 15-items drawn from the pool of 60 items on the BNT. The aim of creating this modified version of the test was to decrease cultural bias and to ensure equivalent performance across different languages. The 60 BNT items were divided into 15 sets of four items of equivalent difficulty; for example, items 1, 2, 3, and 4 of the BNT formed a pool of items and from that pool, item 1 of the modified test was selected. This procedure ensures the items in the short form are of increasing difficulty equivalent to the original test (Mack et al., 1992).

Fifteen South African neuropsychologists, all of whom used the BNT in their clinical practice, made the selection of items for the BNT-SA-SF. These individuals rated each item in the 15 pools according to whether they were culturally appropriate, and then selected the most culturally appropriate of the 4 items in each pool. For instance, the item *beaver* was one of the four options in the eighth pool. However, a beaver is likely to be an unfamiliar animal to individuals in the Western Cape, so *rhinoceros* was chosen. Finally, a team of linguists translated and backtranslated the modified test from English into the other two languages most widely spoken in the Western Cape (Afrikaans and isiXhosa).

The BNT-SA-SF was not administered as a separate measure to participants. Instead, the performance of each participant on the 15-items of the test, was drawn from their performance on the full BNT in order to obtain a score

Procedure. All participants signed up for an individual test appointment via the online SRPP service. At this appointment, I ensured the participant read, understood, and signed the consent form, which informed them about the details of the study and the fact that they could withdraw at any time. They also completed the demographic questionnaire, which was then used to screen potential participants in line with the eligibility criteria listed above. Those not qualifying for the study (n = 7) were excused immediately, but did receive SRPP points commensurate with their amount of participation.

Those meeting the eligibility criteria were administered the BNT. At the conclusion of the test administration, an appointment was set up for exactly 2 weeks later. At the second testing session, participants were only administered the BNT. After this second test administration, participants were debriefed and given the opportunity to ask any questions relating to their testing experience. Approval for all study procedures was obtained from the Research Ethics Committee of the UCT Department of Psychology.

Statistical Analysis. Data were analysed using the SPSS software package, version 18.0. The threshold for statistical significance (α) was set at 0.05, and appropriate effect size estimates were calculated for each analysis. The initial analytic step involved the generation of descriptive statistics for all relevant variables and examination of the assumptions underlying the necessary statistical tests. The relevant parametric tests were used on continuous variables, and chi-square analyses were used on categorical data where assumptions were upheld, unless otherwise stated. For data not meeting the necessary assumptions, non-parametric equivalent tests were used. There were several components to the analytic strategy.

Step 1. I calculated the test-retest reliability, which was estimated using Spearman's Rho, and the internal consistency, which was estimated using Cronbach's α, of the current sample's BNT data. These data were then compared to similar measures of North American normative samples.

Step 2. I compared the mean BNT score of the current sample at test time 1 to North American norms for similar age groups. In addition, I calculated a difficulty index for each item; this index is the proportion of correct responses produced spontaneously or following the presentation of a semantic cue. The difficulty index for each item for the current sample was then compared to similar data from North American samples to help identify items that may be particularly problematic.

Step 3. The analyses here focused on the BNT-SA-SF. Reliability estimates for this group were calculated to provide provisional estimates of the functioning of this test in a South African demographic. I investigated item-by-item performance on the modified short form by creating an item difficulty index for each item, and attempting to detect linear trends in performance via graphically.

Results

BNT.

Reliability analyses. Spearman's rho correlation coefficient was used to estimate test-retest reliability between test time 1 (T1) and test time 2 (T2). This coefficient, rather than Pearson's product-moment correlation coefficient, was used because test scores were non-normally distributed. The analysis showed that although performance at T1 was significantly related to that at T2, the correlation was somewhat lower than might have been expected based on past literature, $r_s = .574$, $p < .001$.

To estimate the internal consistency of the test, I calculated Cronbach's *α* based on performance at T1. The value of this coefficient was within the range most commonly cited as an acceptable value for this statistic, α = 738. Component variables with zero variances were not included in this reliability analysis (see Appendix C).

Overall score: Comparisons to American norms. Table 2 presents demographic and BNT descriptive statistics for North America normative samples and for the current sample. Farmer (1990) was chosen as it one of only a small number of studies containing norms for young adults (BNT normative data is typically presented for older adult age groups). The study by Tombaugh and Hubley (1997), which provides Canadian norms, is also included as it provides BNT norms that are widely used. In addition, Mitrushina et al. (2005)'s meta-analytic norms for the BNT, stratified by age, are included for comparative purposes.

A series of one-sample *t*-tests was used to compare the performance of the current sample with each of the normative groups. Bonferroni correction was employed to decrease the Type I error rate. Thus, *α* was set at .0167 for these comparisons. The current sample scored significantly lower than the Farmer (1990) sample, $t(29) = -3.681$, $p < .001$, the Tombaugh and Hubley (1997) sample, $t(29) = -3.598$, $p < .001$, and the Mitrushina et al. (2005) meta-analytic norms, $t(29) = -2.996$, $p = .006$.

Table 2
Demographic Characteristics and BNT scores for North American Normative Samples and the Current Sample

Study	*n*	Age		Education (years)	Score
		Range	*M* (*SD*)		
Farmer (1990)	25	20-29	24.08 (2.53)	14.88 (1.67)	56.04 (3.60)
Tombaugh and Hubley (1997)	22	25-34	47.5 (12.4)	-	56.00 (2.90)
Mitrushina et al. (2005)	-	25-29	-	-	55.71 (3.08)
Current sample	30	18-25	21.08 (1.89)	13.00 (1.29)	54.27 (2.64)

Note. In columns 4,5, and 6 means are presented with standard deviations in parentheses.

These results seem to indicate that South African English-speaking high functioning adults perform, on average, more poorly than North American adults of similar age. However, these results must be interpreted with caution, as BNT scores were significantly non-normally distributed: Both p-p plots and the Shapiro-Wilk test of normality, *S-W*(60) =

.941, $p = .006$ at T1, $S\text{-}W(60) = .801$, $p < .001$ at T2 confirmed that the distribution was negatively skewed[1].

Item analysis. To further investigate the performance of current sample, a difficulty index for each item was calculated (see Table 3). The desired trend is for the proportion of correct responses to decrease as the test progresses. Thus, all participants should answer the first few items correctly but fewer participants should produce a correct response on the less familiar items towards the end of the test. Those items that appeared significantly more difficult for the South African than the Tombaugh and Hubley (1997) sample are presented in the table. A number of these items, including *beaver* and *hammock,* were on the Mack SF-4, but the experts regarded these items as biased as thus they were replaced in the BNT-SA-SF.

Table 3.
Difficulty Index for BNT Items for Current Sample and the Tombaugh and Hubley sample.

Item	Current sample *n* = 30	Tombaugh and Hubley sample *n* = 219
26. canoe	93.33	100.0
28. wreath	53.33	99.5
29. beaver	90.0	97.5
30. harmonica	80.0	96.8
39. hammock	83.33	94.1
40. knocker	90.0	97.7
42. stethoscope	90.0	95.0
44. muzzle	83.33	92.7
47. accordion	50.0	81.7
48. noose	80.0	91.3
54. tongs	56.66	84.5
56. yoke	26.67	63.0
57. trellis	26.67	77.2

Note. The items included in this table are those items for which there was a difference of 5% or more between the two samples.

BNT-SA-SF.

Reliability analyses. Table 4 presents statistics describing the performance of the current sample on the BNT-SA-SF. Examination of the p-plots and box-and-whisker plots of these data revealed that they were non-normally distributed. Although log and square-root transformations were attempted, neither improved the normality of the distributions. Therefore, Spearman's correlation coefficient is reported as an estimate of test-retest reliability, $r_s = .390$, $p = .033$. To investigate whether scores from the short form agreed with

[1] Log and square route transformations were attempted but neither improved the distribution of BNT score.

scores from the full test, the correlation coefficient of the BNT-SA-SF with the BNT was calculated. There was a significant positive association, $r_s = .536$, $p = .002$ at T1, but not at T2, $r_s = .200$, $p = .288$

With regard to internal consistency of the modified test, there were too few items with variances > 0 to compute Cronbach's α, as all participants answered the first ten items correctly.

Table 4.
Performance by the Current Sample on the BNT-SA-SF

	Test time 1 (*n* = 30)	Test time 2 (*n* = 30)
Range	12-15	13-15
M (*SD*)	14.3 (0.75)	14.8 (0.48)

Item analysis. Figure 1 presents the difficulty index for each of the 15 items on the modified test. The trend for increasing errors as the test progresses is evident, with the all participants getting the first 10 items correct, and participants making increasingly more errors on items 13-15. Item 12, *funnel,* seemed to be a slightly more familiar item for this sample than item 11, *stethoscope.*

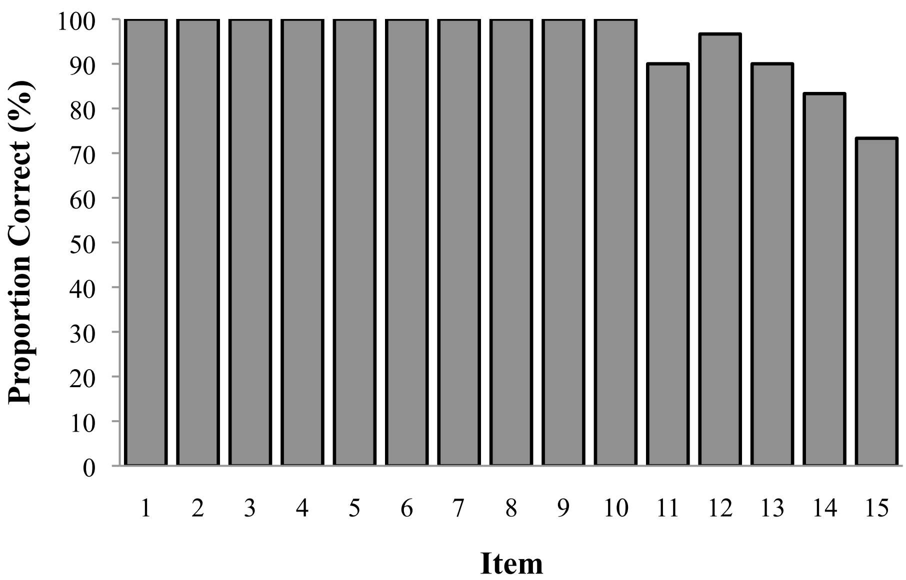

Figure 1. Proportion of correct responses made spontaneously or with stimulus cue for South African English-speaking young adults on the BNT-SA-SF.

Discussion

Investigation of the performance of a group of English-speaking South African undergraduate students with a high level and quality of education and from high SES backgrounds indicated that the BNT, in its standard form, was a fairly reliable instrument within this population group. Although a number of items were identified as problematic, such as *wreath,* items that were expected to be problematic according to other cross-cultural studies of the BNT (Barker-Collo, 2001, 2007), such as *beaver* and *pretzel*, were not problematic in this sample. This finding suggests that this South African sample is probably more familiar with North American culture than other non-North American samples are. However, direct comparisons of the average performance of the current sample to North American standardization data showed that the sample performed slightly below age-matched norms on overall score and on the reliability measures. These data, then, suggest that there is still an element of cross-cultural bias in favour of North American samples, particularly considering the demographic sample hypothesised to perform optimally still did not meet North American performance levels.

With regards to the internal consistency (.738) of the BNT in the current sample, previous literature reports slightly higher reliability coefficients (between .78 and .96) for the test (Fastenau et al., 1998; Grave et al., 2004). This finding seems to suggest that some participants were more familiar with particular test items than others. Similarly, previous studies also report higher short-term test-retest reliability (at about r = .91, with a gain of about one point on the second test session) than was found in the current study. These previous estimates were taken from healthy older adult samples, though (Strauss et al., 2006). Reliability data for young adults could not be located. This finding could indicate the unreliability of test items to students, in that some students were not familiar with test items at T1, but got these items correct at T2 i.e. practice effects. In addition, a ceiling effect could have influenced the low test-retest reliability, in that those who scored highly i.e. 15 out of 15, at T1 could not improve their score at T2.

The current study also sought to investigate the functioning of the BNT-SA-SF, in the same sample of English-speaking South African undergraduate students. Performance on the BNT-SA-SF was positively correlated (r = .54) with performance on BNT. The value of this correlation coefficient is lower than that reported for other short forms with the standard form; correlation coefficients reported in the literature on other 15-item versions range from

.97 (Mack et al., 1992) to .63 (Fastenau et al., 1998). A possible reason for this discrepancy between the current findings and those of previous studies is that, because the BNT-SA-SF specifically sought to remove problematic items that are part of the BNT, (i.e., items such as *hammock* and *wreath*, on which the South African participants scored poorly), the average scores were higher on the modified short form than on the BNT. This may therefore be an indication that the BNT-SA-SF is less biased against South African samples than the standard BNT, and is a promising result.

Test-retest reliability of similar 15-item versions has been reported to be as high as .90 (Fastenau et al., 1998). In the current sample, test-retest reliability was fairly low, at r = .39. Once again, these data are indicative of the fact that the sample performed well on the BNT-SF-SA, with a mean score of 14.3 at T1. Therefore, a ceiling effect would have come into play for those participants scoring 14 or 15 at T2.

Overall, the reliability analysis of the BNT-SA-SF showed that young South African adults who are English-speaking, of high SES, and with a high quality and level of education perform relatively better on this modified test than on the standard BNT. Furthermore, the analysis identified no problematic items in the BNT-SA-SF, suggesting that appropriate items had been selected for the test.

Study 2

Study 2 aimed to investigate the validity of the BNT-SA-SF. The cross-cultural validity of the test was investigated by examining whether a group of isiXhosa-speaking individuals of low SES performed differently from a group of English-speaking individuals of high SES. The clinical validity of the test was assessed by examining whether (1) there was a significant difference in performance of clinical patients on the Mack SF-4 short form (Mack et al., 1992) compared to the modified test, and (2) either the Mack SF-4 or BNT-SA-SF distinguished South African patients with AD from those with VaD.

Methods

Design and setting. A quantitative, non-equivalent groups design with historical controls was used in accordance with nonrandomized selection criteria. There was a single administration of the BNT and Mini Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975) to newly-recruited participants.

To determine the cross-cultural validity of the BNT-SA-SF, community-dwelling healthy older adults from high and low SES backgrounds and with different home languages were recruited from various Cape Town communities.

To determine the diagnostic efficiency of the BNT-SA-SF, data were collected from patients referred to the UCT's Groote Schuur Memory Clinic, a programme of the Albertina and Walter Sisulu Institute of Aging in Africa (IAA). The clinic serves to diagnose and provide treatment options for older adults with various forms of dementia. Patients of varying race, SES, education and language present at the Memory Clinic; most are older than 50 years of age. Patients who present here have been referred because of noticeable changes in or problems with their memory functioning (Kalula et al., 2010). As part of giving their consent to receive clinical services at the clinic, all patients give their permission for clinic-affiliated researchers to use data derived from their assessment.

Participants. The healthy older adult sample was comprised of 60 community-dwelling adults. The first group (English OA; *n* = 30) included older adults who had English as a home language, who emerged from high-SES backgrounds, and who had received a high level and quality of education. They were recruited from various Cape Town communities. The second group (Xhosa OA; *n* =30) included older adults who had isiXhosa as a home language, who emerged from low-SES backgrounds, and who had received a low level and quality of education. They were recruited from a community organization (Grandmothers Against Poverty and AIDS; GAPA) located in Khayelitsha. Ethical approval for the testing of these two groups was obtained from the Research Ethics Committee of the UCT Department of Psychology.

Table 5 presents demographic characteristics for each healthy OA group. As can be seen, there were no statistically significant between-group differences concerning age. As expected given the aim of the study and our recruitment strategy, there were significant between-group differences with regard to education level and SES. Additionally, there were no males in the Xhosa OA group, whereas the English OA group had equal numbers of males and females. The majority of recent research findings do not report significant differences between male and female older adults (Kent & Luszcz, 2002; Saxton et al., 2000; Zec et al., 2007b).

Table 5
Demographic Information for the Two Healthy Older Adult Groups

	Group				
	English OA (*n* = 30)	isiXhosa OA (*n* = 23)	Test Statistic	*p*	ESE
Age (years)[a]					
Range	60.42-74.17	50.25-71.58			
M (*SD*)	65.69 (3.41)	63.71 (5.87)	1.422[b]	.159	.045
Sex (M:F)	15:15	0:23	18.009[c]	<.001***	.550
Race (White:Black)	30:0	23:0	53.000[c]	<.001***	1.00
Education (years)					
Range	12-23	5-12			
M (*SD*)	15.2 (2.89)	8.78(2.26)	8.780[d]	<.001***	.602
SES (Hi:Med:Lo)[e]	25:5:0	0:0:23	63.271[c]	<.001***	1.00

Note. The test statistic is either *t* or χ^2. ESE = effect size estimate.
[a] Participant age, represented in years, was calculated from participant's age in months at time of testing, which was divided by 12 to achieve a decimal value.
[b] df_{within} = 21.094 (Levene's test returned a significant result; therefore the reported statistics are for equal variances not assumed).
[c]The statistics reported are based on Fisher's exact test because the data did not meet the assumptions for a chi-square analysis (1 or more expected counts were < 5).
[d]df_{within} = 51
[e]SES was estimated using 2001 census data for the median annual income level for each participant's suburb of residence. High, medium, and low SES brackets were defined according to the 1st and 3rd quartiles such that, for instance, those median annual suburb incomes that fell below the 1st quartile were classified as low SES.

The clinical sample was comprised of 48 patients who had presented to the Memory Clinic in 2009 and 2010. This sample was split into two groups. The first group (Old BNT; *n* = 24) was comprised of new patient admissions to the clinic in 2009, who were ultimately diagnosed with AD or vascular dementia, and who were administered the Mack SF-4[2]. The second group (New BNT; n = 24) was comprised of new patient admissions to the clinic from January to September 2010, who were ultimately diagnosed with AD or vascular dementia, and who were administered the BNT-SA-SF.

Table 6 presents demographic characteristics for each clinical group. As can be seen, there were no between-group differences with regard to sex distribution, race, age, level of education, or MMSE score.

[2] Prior to 2009, no Boston Naming Test form was used as part of the Memory Clinic's neuropsychological test battery.

Table 6

Demographic Characteristics of Memory Clinic Patients

Variable	Old BNT AD (*n* = 16)	Old BNT VaD (*n* = 8)	New BNT AD (*n* = 17)	New BNT VaD (*n* = 7)	Test statistic (F/χ^2)	*p*	ESE
Sex (M:F)	6:10	4:4	4:13	1:6	2.865	.423	.407[a]
Race (Coloured:White)	8:8	6:2	11:6	6:1	2.516	.405	.260[a]
Age (years)							
Range	50-86	49-80	49-84	58-81			
M (*SD*)	71.89 (10.28)	68.88 (11.73)	71.76 (9.86)	73.29 (7.76)	0.258[a]	.855	.017[c]
Education (grades)					8.339[d]	.214	.295[a]
≤ 7	2	1	3	3			
8-11	6	5	10	4			
≥12	8	2	4	0			
MMSE							
Range	12-30	13-30	9-27	13-28	0.791[a]	.505	.051[c]
M (*SD*)	22.06 (5.26)	22.38 (6.30)	19.76 (4.96)	19.86 (5.46)			

Note. The Old BNT group was administered the Mack SF-4 (Mack et al., 1992). The New BNT group was administered the instrument being evaluated in this study, the BNT-SA-SF. AD = Alzheimer's disease. VaD = vascular dementia. ESE = effect size estimate.

[a]Effect size estimated using Cramer's V.

[b]df_{within} =44.

[c]Effect size estimated using partial η^2.

[d]Fisher's exact test (α_e) is reported as the data did not meet the assumptions of a chi square analysis (1 or more expected counts were <5).

Inclusion and exclusion criteria. For the healthy adult participants, an MMSE score < 24, the most widely used cut-off score for cognitive impairment on this measure of global cognition, warranted exclusion from the study to ensure performance on the BNT was not confounded by cognitive impairment. Based on this criterion, seven participants from the original Xhosa OA sample had to be excluded. Thus, the final sample size for Xhosa OA group was 23. No participants were excluded from the English OA group.

To ensure participants were healthy, an extensive self-report medical history was obtained from each individual. The same criteria as in study 1 were applied here.

For the clinical participants, only patients who had been given a primary diagnosis of possible or probable AD, mixed dementia, or VaD were included; all alternate primary diagnoses were reasons for exclusion. Further, of these patients, only those for which there were full and accurate records of MMSE scores and demographic information, and who were administered the full 15 items of a short-form BNT, were included. This criterion was particularly important in gathering data for the 2009 sample from the clinic archives.

Measures. All participants were administered the MMSE and a BNT short form (either the Mack SF-4 or BNT-SA-SF). The MMSE, a measure of global cognition with lower scores indicating greater general cognitive impairment, is a widely used test that is frequently used to ascertain a measure of general functioning in older populations (Folstein et al., 1975). The test was administered in the language of the participant's choice. An isiXhosa translation of the MMSE, which had been used in previous published research, was obtained for this purpose (Roos et al., 2010) and all tests for isiXhosa-speaking participants were administered by a first-language isiXhosa-speaking graduate.

Procedure. Participants in the English OA group were tested individually in a quiet room in their home of residence. Participants in the Xhosa OA group were tested individually in a quiet room at the GAPA premises.

Each participant in the two groups completed a consent form (see Appendices A) and demographic questionnaire (see Appendix B for the English version) before being tested. The latter was used to screen participants according to the inclusion and exclusion criteria; those who did not meet eligibility criteria were not tested further. Following this introductory part of the test session, the examiner administered the MMSE and the BNT-SA-SF (see Appendix C) in the participant's home language. The examiner gave clear instructions before each test was administered.

After test administration concluded, participants were fully debriefed and given the opportunity to ask questions or express opinions about the test procedure. Each participant in the Xhosa OA group received a sum of R30 to cover their travel costs to and from the GAPA centre.

Participants in the clinical sample were tested as part of their neuropsychological assessment session during their appointment at the Memory Clinic. On presentation at the clinic, a team of professionals, including a medical doctor, neuropsychologist, and psychiatrist, examine the patient and come to a collaborative diagnosis. The patient is administered a standard screening battery of neuropsychological tests, of which the MMSE and BNT short form are two tests, as one facet of this examination.

Statistical Analysis. These analyses were conducted to assess the cross-cultural and diagnostic validity of the BNT-SA-SF. Data were analysed using the SPSS software package, version 18.0. The threshold for statistical significance (α) was set at 0.05, as it was in study 1, and the same initial analytic steps as were conducted in study 1 were conducted here. All assumptions for the necessary tests were met, unless otherwise stated.

Cross-cultural validity. To assess whether the BNT-SA-SF was of equivalent difficulty across the English OA and Xhosa OA groups, an independent samples *t*-test was carried out. Following this, I performed a multiple regression analysis, predicting BNT score from demographic variables of SES (defined as high, middle, or low according to the mean annual income of the participants' suburb of residence) and education (measured in grades completed). Lastly, I carried out an item analysis to identify patterns of performance of the two groups across the 15 test items.

Diagnostic validity. Assessing the clinical validity of the BNT-SA-SF involved a bbetween-group comparison using a 2(group) x 2(diagnosis) factorial analysis of variance (ANOVA). I used ANOVA to determine whether (a) performance of the New BNT group differed significantly from performance of the Old BNT group, irrespective of diagnosis, (b) those patients diagnosed with AD scored significantly lower than those diagnosed with VaD, irrespective of test version, and (c) whether there was an interaction effect between test version and diagnosis (for example, whether AD patients scored significantly more poorly than VaD patients on the BNT-SA-SF but not on Mack SF-4). I concluded the analysis by using correlation techniques to investigate the relationship between test score and the demographic characteristics of participants. Correlation was used, as opposed to regression, due to the small sample size and the fact that a minimum of 10 – 15 cases per predictor is recommended (Field, 2009).

Results

Cross-cultural validity. As shown in Table 7, an independent samples *t*-test indicated that healthy English-speaking OAs performed significantly better than healthy isiXhosa-speaking OAs on the BNT-SA-SF.

Table 7
Results of Between-Groups Comparison of BNT-SA-SF performance

	Group				
	English OA	isiXhosa OA	$t(51)$	*P*	ESE
Range	11-15	5-10			
M (*SD*)	14.10 (1.06)	7.61 (1.37)	19.421[a]	< .001***	.881

Note. ESE = effect size estimate. Partial η^2 is reported here.
***$p < .001$

Hierarchical regression. I performed further analysis to investigate the contribution of SES and education to BNT-SA-SF performance in these groups. The regression model was created by first entering SES (measured by the method described in the demographics table above) and education (measured by years of completed education) together as predictors in one block; this step is appropriate as these two measures are related in this sample. The isiXhosa OA group is defined by low SES and low level of education (≤12 years) and the English OA group is defined by medium/high SES and higher levels of education (≥12 years). Age and sex were entered at the second and third steps respectively. There is consistent evidence in the literature that age is associated with BNT performance but there is little substantial evidence for the influence of sex on BNT performance, with most studies not finding a significant effect (Kent & Luszcz, 2002; Tombaugh & Hubley, 1997). However, as the groups differed on this variable, it was important to assess whether this was associated with BNT performance.

The results of the regression model are shown in Table 8. As can be seen, SES and education level together had a significant effect on BNT performance. Neither sex nor age was a significant predictor of BNT performance in this sample, however. The accuracy of the regression model was checked using diagnostic tests. There were no problems with collinearity or outliers, VIF were all < 10 and tolerance statistics were all > .35. In addition, the residuals were roughly normally distributed. However, examination of the plots of standardized residuals indicated education and sex may violate the assumption of heteroscedascity in this sample, and therefore generalizing this model beyond the current sample is cautioned. This caution is acceptable, considering this is a preliminary analysis.

Table 8
Regression Analysis of BNT-SA-SF Score

	B	SE *B*	*β*	*t*	*p*
Step 1					
Constant	14.824	1.304		11.368	< .001***
Education	0.099	0.071	0.118	1.396	.169
SES	-1.989	0.203	0.831		< .001***
Step 2					
Constant	16.285	1.532		10.629	< .001***
Education	0.078	0.070	0.093	1.106	.274
SES	-1.899	0.205	-0.793	-9.248	< .001***
Age	-0.826	0.477	-0.109	-1.731	.090
Step 3					
Constant	17.834	3.535		5.319	< .001***
Education	0.070	0.073	0.084	0.962	.341
SES	-1.932	0.216	-0.807	-8.931	< .001***
Age	-0.821	0.481	-0.108	-1.707	.094
Sex	-0.002	0.003	-0.029	-.521	.605

Note. R^2= .856 (p <.001) for Step 1; ΔR^2 = .008 (p = .090) for Step 2; ΔR^2 = .001 (p = .605) for Step 3.
***p < .001

Item analysis. Figures 2 and 3 present difficulty indexes for each BNT item for the Xhosa OA and English OA groups respectively. The difficulty index is calculated and interpreted in the same manner as study 1.

The desired trend can be detected in Figure 2, with a gradual decrease in the proportion of participants producing a correct response as the test progresses, except for item 14 (*sphinx*) on which more participants seemed to produce a correct response than on the previous item. Figure 3 reveals a very different trend. Most participants were able to produce a correct response on the first six items; however, it is evident that very few participants were able to name items 7, 8, 9, 13, or 15 correctly, with no participants producing a correct response on items 11 and 14. Items 10 and 12 appear to be have been easier for this sample than the adjacent items were.

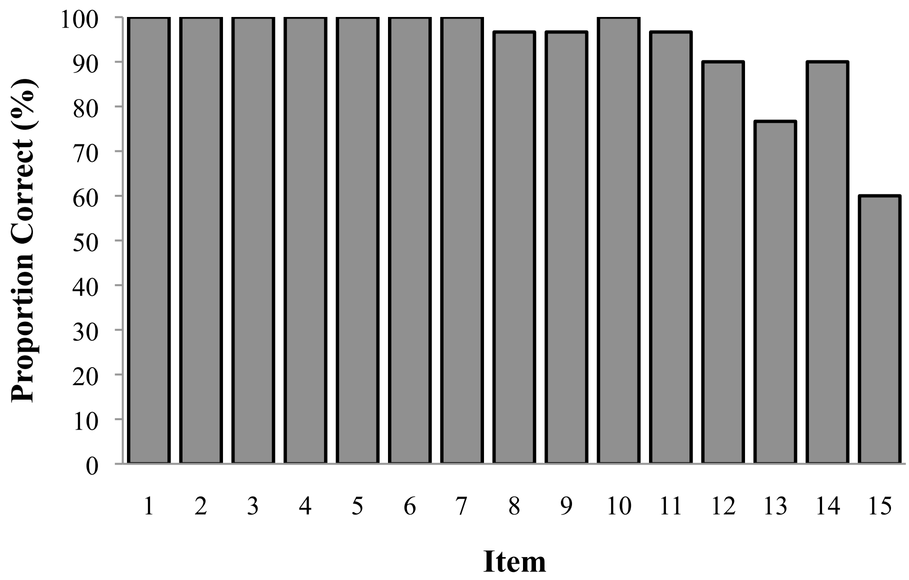

Figure 2. Proportion of correct responses made spontaneously or with stimulus cue for English OA group.

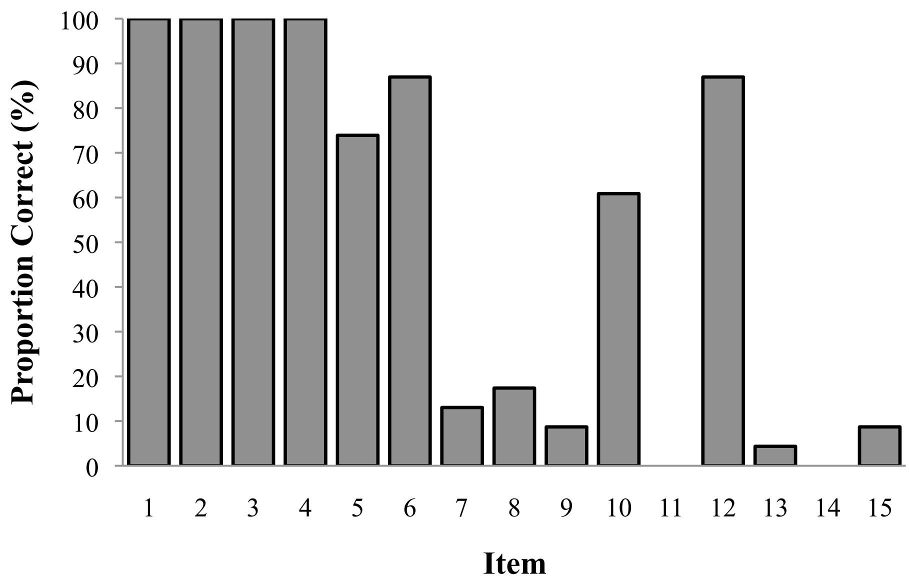

Figure 3. Proportion of correct responses made spontaneously or with stimulus cue for Xhosa OA group.

Clinical validity

I explored the diagnostic ability of the BNT-SA-SF compared to the original Mack SF-4 by investigating how well the tests were able to distinguish between patients diagnosed with dementia of the AD and those diagnosed with VaD. The expectation here was that overall patients in the New BNT group would score higher than patients in the Old BNT group as the original test was biased against the South African population, meaning that patients scored lower as indication of unfamiliarity with items, not of impairment. For this same reason, it was expected that there would be a greater difference between AD and VaD in the New BNT group than in the Old BNT group.

Between-group comparisons. I used ANOVA to investigate the hypothesis that, in patients administered the BNT-SA-SF, those diagnosed with AD would score significantly lower than VaD patients would, but that, in patients administered the Mack SF-4, those diagnosed with AD would not score significantly differently from those diagnosed with VaD patients. More specifically, I conducted a 2 (group: Old BNT vs. New BNT) x 2 (diagnosis: AD vs. VaD) factorial ANOVA with BNT score as the dependent variable. Initial analysis of the distribution of BNT score revealed that it deviated slightly from normality, however ANOVA is robust to small deviations from normality, especially when other assumptions are upheld (Field, 2009).

Table 9 presents descriptive statistics for the data used in this analysis. The ANOVA did not detect a statistically significant effect for either of the independent variables, diagnosis, $F(1, 44) = .262$, $p = .611$, partial $\eta^2=.006$, or BNT version, $F(1, 44) = .215$, $p = .645$, partial $\eta^2=.005$. Therefore, BNT score was not significantly different across the two versions of the test or between patients with AD and VaD. Further, there was no statistically significant diagnostic group by test version interaction effect, $F(1,44) = .001$, $p = .973$, partial $\eta^2 = .001$, suggesting that, using this analytic technique, neither test was able to distinguish between patients with AD or VaD.

Table 9
Descriptive Statistics of BNT Score on the Original and Modified Tests for the Clinical Sample

	Diagnosis		
Test Version	AD	VaD	AD+VaD
Mack SF-4	9.25 (3.80)	9.75 (3.15)	9.42 (3.54)
BNT-SF-SA	9.71 (1.90)	10.14 (2.41)	9.83 (2.01)
Both Test Versions	9.48 (2.94)	9.93 (2.74)	9.63 (2.86)

Note. Means are presented with standard deviations in parentheses.

Although the ANOVA failed to return a significant result, examination of the resulting means plot (see Figure 4) shows evidence of an interesting trend in the data. Overall, patients achieved higher mean scores on the modified test than on the original test. Further, VaD patients achieved higher mean scores than VaD patients across both tests. Therefore, a power analysis was conducted to estimate the power of the current sample size and estimate the sample size required to detect a significant difference between AD and VaD patients on the BNT-SA-SF. The power of the current sample size (*n* = 24) to detect such a difference is 0.122. The sample size needed for a significant 1-tailed *t*-statistic, with a power of .80, is 614 (when α is set at .05 and the effect size is small, *r* = .10).

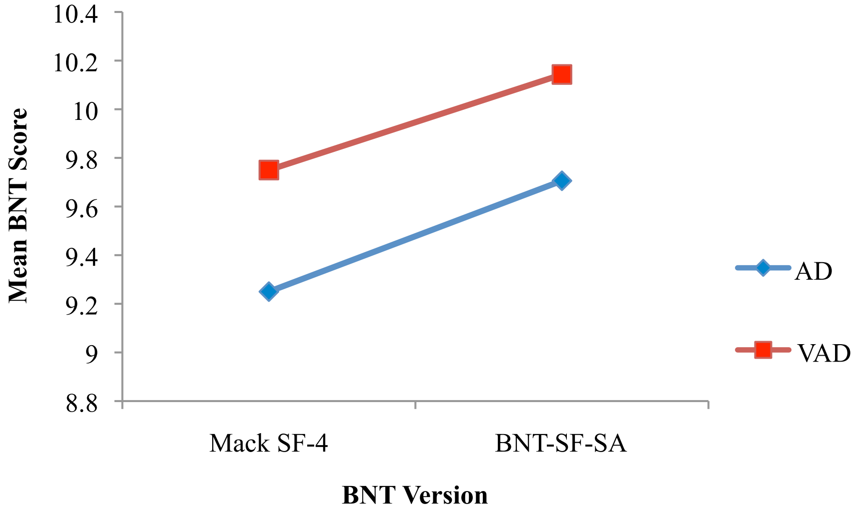

Figure 4. Means plot showing mean BNT score for AD and VaD patients on the Mack SF-4 and the BNT-SA-SF.

Correlation analyses. These analyses were performed to ascertain whether diagnosis, demographic factors, and MMSE score were related to each other, and more importantly, to performance on the BNT-SA-SF. Table 10 presents Pearson's correlation coefficients on these variables for participants who were administered the BNT-SA-SF. As can be seen, performance on the modified test is statistically significantly related to sex (with males scoring higher than females). Furthermore, sex and MMSE score are significantly related to one another (with males scoring higher than females), as are education and race (with white patients falling in a higher education bracket than coloured patients). Interestingly, however, diagnosis is not significantly correlated to test score, which supports the above finding that the test does not discriminate strongly between patients diagnosed with AD and those diagnosed with VaD.

Table 10
Relationships between BNT-SF-SA Performance, Demographic Variables, and MMSE Score.

	1	2	3	4	5	6	7
1. BNT-SF-SA score	1.000	.101	.077	**-.512***	-.178	.187	**.503***
2. Diagnosis		1.000	-.077	-.103	-.210	-.346	.009
3. Age			1.000	-.125	.292	.250	-.129
4. Sex				1.000	.103	-.387	**-.442***
5. Race					1.000	**.513***	-.292
6. Education						1.000	.221
7. MMSE score							1.000

Note. All reported correlations are Pearson's correlation coefficient (r).
*$p < .05$

Discussion

Cross-cultural validity. The finding that the Xhosa OA group scored significantly lower than the English OA group suggests that the modified test does contain cultural bias. Regression analyses confirmed this impression, indicating that education and SES differences, rather than age and sex differences, accounted for the significantly lower scores within the Xhosa OA group. Mosdell et al., (2010) reported a similar result, with English speakers' average scores on a 30-item BNT significantly higher than those of isiXhosa speakers.

A possible explanation for the poor performance of the isiXhosa group is that the sample was homogenous, due to the nature of the organisation from which these participants were recruited, and their rural backgrounds. They all shared a similar lived experience, in terms of having lost a child or grandchild to HIV/AIDS. Further, many attended school in similar rural areas of the Transkei. As Shuttleworth-Jordan (1996) notes, individuals more integrated with mainstream, westernized, urban culture, irrespective or race and language, are likely to perform reasonably well (i.e., at levels much closer to those of Western standardization samples) on neuropsychological tests. These isiXhosa-speaking older adults, almost all of whom were from rural backgrounds, were perhaps less familiar with particular BNT items than other isiXhosa-speaking low-SES individuals (for instance, urbanized young adults) might be.

A limitation of this study is that group, education, language, and SES were intercorrelated. In other words, all the participants in the isiXhosa group were of low SES and had a fairly low level of education, whereas all the participants in the English group were of high SES and had a fairly high level of education. In addition, the sample sizes were fairly small from which to draw generalizable conclusions. Lastly, although gender effects are not expected, as the literature tends to indicate non-significant effects for gender (Kent & Luszcz, 2002; Zec et al., 2007a) it would be valuable to include male participants in further studies. For these reasons, it needs to be investigated whether similar patterns of responses are reproduced when the test is administered to a broader sample of isiXhosa-speaking low SES individuals. Future research using into the BNT-SA-SF should investigate how English- and isiXhosa-speaking adults of varying SES and levels of education perform on the test. This will provide evidence as to whether the between-group differences found in this study are consistent across all levels of SES and education.

Although further research with larger and more diverse samples is necessary, the preliminary results reported here indicate that a relatively poor BNT-SA-SF score is not necessarily indicative of a serious underlying neurological impairment. It is important for researchers and clinicians to be aware of the significant effect demographic variables may have on BNT performance, even when the test has been modified for use in South Africa. Furthermore, these preliminary results seem to confirm that gathering separate normative data for different sectors of the South African population might be a more efficient endeavour than attempting to create a BNT version that performs equally across different language and SES groups.

Diagnostic validity. The mean BNT score of AD patients was lower than that of VaD patients on both the Mack SF-4 and the BNT-SA-SF. The finding that this difference was not statistically significant is not surprising considering the difficulty in differentiating between AD and VaD (Mathias & Burker, 2009). Such diagnosis usually involves combined assessment across multiple tests and other sources of information, as opposed to relying on a single test (Lukatela et al., 1998). Thus, the small differences in scores that emerged in this study do not necessarily indicate that the test does not have clinical utility, but rather, that it must be used in tandem with other diagnostic measures. In addition, there is little evidence that short forms of the BNT discriminate as well as the standard 60-item version in this regard.

There is more evidence, however, that short forms do discriminate as well as the long form between normal healthy persons and those with AD (Larrain & Cimino, 1998). A suggestion for future research on the BNT-SA-SF is therefore to compare the performance of AD patients either alone or in combination with VaD patients, to healthy controls. A comparison of the mean BNT-SF-SA scores in this study shows that English OA group was at 14.10 and the memory clinic patients at 9.92. That is a promising result, but excitement must be tempered by two facts: First, the memory clinic participants are drawn from a different demographic sector of the South African population (they tend to be of lower SES and with lower levels of education), and, second, the isiXhosa-speaking group achieved an average score of 7.61 on the BNT-SA-SF. Therefore, if demographic characteristics are not taken into consideration, one might diagnose cognitive impairment where none exists.

General Discussion

The BNT is one of the most widely used tests of confrontation naming ability. The enduring popularity of the test is evident in the creation of various 15- and 30-item short forms and the modification of the test for use within different cultural and language groups. The inherent bias in the test to those not familiar with North American culture, and the need for shorter tests to facilitate shorter test times in clinical settings, led researchers from the Universities of Cape Town and Stellenbosch to develop a 15-item version, thought to be more culturally appropriate than the standard 60-item test or existing short forms developed in other countries.

In analysing the reliability and validity of this modified test, it is clear that the BNT-SA-SF has not solved the cultural bias issues, and that performance on the test is heavily influenced by demographic variables. Specifically, the preliminary results reported here indicate that healthy individuals with limited education and low SES perform more poorly than age-matched healthy individuals with higher levels and quality of education, and from higher SES backgrounds.

A limitation of this study was the small sample sizes, particularly with the clinical sample in study 2. Future research on the BNT-SA-SF should look to include a much larger sample across both clinical and control participants. Nevertheless, these findings mirror the conclusions drawn in the extant literature, and by practicing clinicians, that it is vital to consider demographic variables such as SES and education when assessing patient performance on the BNT (Shuttleworth-Edwards et al., 2004; Strauss et al., 2006). More importantly, perhaps, subjects with low levels of education, of varying SES, and of different languages must be included in the creation of norms (Hawkins & Bender, 2002; Uzzell, Ponton, & Ardila, 2007).

Nonetheless, the results of this research provide valuable information for South African clinicians and researchers who utilise the BNT, in terms of extraneous variables that may influence individual performance on the test and in terms of confirming clinical lore that a poor score on the test is not necessarily an indication of serious underlying cognitive impairment.

Further, 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. This 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 purchasing 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. Another advantage of this short form is that, unlike many other short forms that are developed via odd-even or split -halves methods, this one was developed on an item-by-item basis, which lends itself to evaluation by item response theory (Pedraza et al., 2009).

Two broad suggestions for the way forward are these:

(1) 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) new items, of equivalent difficulty. Alternatively, more items could be added to the test, as longer versions of the test have been found to be more discriminative in clinical populations than short versions.

    One way to go about this change, which other authors have successfully used (e.g., Saxton et al., 2000), is to administer the full 60-item test to a large sample representative of the population, and then to rank the percent correct scores. This would involve the creation of a distribution for each item, as was done in the present study, to create a distribution showing easier to more difficult items in that population, and to then assign items to the new test according to their rank.

(2) To keep the BNT-SA-SF as it is at present and norm this version, providing a large body of data from well-defined community-dwelling older individual. Adopting such a strategy would reduce the potential for misdiagnosis due to the influence of factors such as SES on test performance, and would ideally provide clinicians and researchers with norms stratified by age and education within SES/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. That is, one would have to rearrange the order of the items to ensure that there is a clear linear trend from easier items to those that are more difficult for the South African population.

In closing, it must be emphasised that the samples in this study were drawn from very specific demographic groups in the Western Cape, and that they are therefore not representative of the performance of the general South African population on the various BNT versions used in this study. This research is, however, another valuable step in the ongoing attempt to provide culturally appropriate neuropsychological tests and norms for clinical and research purposes in South Africa. It is also indicative of the difficulty in creating tests that perform equally across the diverse South African population. It is hoped that this preliminary investigation into the reliability and validity of the BNT-SA-SF is explored further, particularly with regard to one or more of the suggestions above.

Appendix C

Items Included/excluded from the Reliability Analysis

Table D1

Items Excluded and Included in Calculation of Cronbach's Alpha.

Variance = 0	Variance > 0
item 1	item 19
item 2	item 26
item 3	item 28
item 4	item 29
item 5	item 30
item 6	item 35
item 7	item 38
item 8	item 39
item 9	item 40
item 10	item 41
item 11	item 42
item 12	item 44
item 14	item 46
item 15	item 47
item 16	item 48
item 17	item 49
item 18	item 50
item 20	item 51
item 21	item 52
item 22	item 53
item 23	item 54
item 24	item 55
item 25	item 56
item 27	item 57
item 31	item 58
item 32	item 59
item 33	item 60
item 34	
item 36	
item 37	
item 43	
item 45	

Note. Items in column 1 were not included in the analysis. Items in column 2 were included in the analysis.