CLAIRE HADDLESEY Bsc (Hons), Msc

Exploration of the impact of language and culture on neuropsychological tests.

Section A: Addressing difference as a result of language and culture when completing neuropsychological tests in a second language: A review of the literature.
Word Count: 8078

Section B: Exploration of the impact of completing the Short Parallel Test of Neuropsychological Status (SPANS) in a second language.
Word Count: 7129

Overall word count: 15207

A thesis submitted in partial fulfilment of the requirements of Canterbury Christ Church University for the degree of Doctor of Clinical Psychology

April 2016

SALOMONS
CANTERBURY CHRIST CHURCH UNIVERSITY

Acknowledgements

I would like to thank all of the people who gave up their valuable time to participate in this study.

I would additionally like to thank my research supervisor Dr Jerry Burgess for his support, knowledge and expertise, both of the SPANS and the research process in general. Jerry's calming nature definitely helped in times of blind panic. My external supervisor, Dr David Eley for his support when I thought that recruitment was a failure, his pragmatic and practical approach got me back on track. To both Jerry and David for their advice and support with my draft manuscripts.

Finally to my family and friends, particularly my fiancée Ryan, whose tireless support through tears and tantrums helped me to survive this at times, daunting process and my friends Kyla and Ron whose efforts to help with recruitment truly saved the day.

Summary of the MRP

Section A provides a critical review of the literature pertaining to language and cultural when completing neuropsychological tests with people whose first language or culture is not the same as the tests. An overview of the studies is provided with methodological critique followed by recommendations for future research.

Section B is a research paper exploring the impact of language and culture on the Short Parallel Test for Neuropsychological Status (SPANS). The SPANS was tested on participants whose first language is not English and compared to a normative sample. Differences between index scores were computed and recommendations for future use with this population made. Additionally, the experiences of being tested in a second language were qualitatively explored.

Section C includes appendices of supporting material for Section's A and B.

Contents

Section A

Addressing difference as a result of language and culture when completing neuropsychological tests in a second language: A review of the literature

Claire Haddlesey

April 2016

Word Count: 8078

A thesis submitted in partial fulfilment of the requirements of Canterbury Christ Church University for the degree of Doctor of Clinical Psychology

Abstract

Background: Neuropsychological tests used routinely within services have a need to be culturally relevant to the populations being tested, as well as consideration for the impact of differences in language.

Methods: This paper presents a critical review of the literature exploring the impact of language and cultural factors on neuropsychological testing.

Results: The review of 14 studies suggests that issues relating to language and culture are prevalent on a number of neuropsychological tests. Cultural familiarity with test-items is a key feature in reducing the differences between groups on a number of tests. The quality of education received was shown to have an impact on tests such as the WAIS-III, with level of education also being demonstrated to be an important factor. Direct translation of tests was shown to be problematic, particularly if care was not taken to control for cultural familiarity and the impact of education and test-wiseness.

Conclusions: There is a need for further research exploring the impact of language and culture on neuropsychological tests used routinely with clinical populations in the UK.

Keywords: Neuropsychological tests, Language, Culture, Culture-fair

Addressing difference as a result of language and culture when completing neuropsychological tests in a second language: A review of the literature.

1. Introduction

1.1 Neuropsychological Tests

Neuropsychological tests are typically used as part of a comprehensive assessment to diagnose neurologically based conditions and to support rehabilitative plans following stroke, acquired/traumatic brain injuries or in the diagnosis of dementia. Additionally, they may be used to test cognitive ability as part of educationally-based assessments or in the diagnosis of learning disabilities (for example Silver, et al. 2006; Silver, et all. 2008). More recently neuropsychological tests have been used within job-related services to assess abilities in relation to unemployment and receiving benefits (Friedli & Stearn, 2015). The tests are typically standardised and have normative data which is usually stratified by age. Neuropsychological tests tend to be developed in Western countries with normative data based on white, English speaking participants (Nell, 2000).

The increased globalisation of economies and widening membership of the European Union (EU) have contributed to an increase in international migration. Within Western Europe and countries with high-incomes, migration has been the main reason for increasing ethnic and cultural diversity (Coleman 2013). Within multicultural societies, research indicates that migrants and their children are more likely to identify with their culture of origin, become bilingual and are less likely to assimilate into their host culture (Bisin & Verdier, 2000; Lezak, 2012). As a result, countries such as the United Kingdom (UK), are seeing increasing numbers of non-native individuals presenting to services for neuropsychological testing. Such individuals may have a different language of origin or culture from the assessor and the available normative sample. There is therefore a need for neuropsychological tests to have normative data or clinical consideration which reflects these more diverse cultural and linguistic populations.

Challenges with neuropsychological testing present when the first language of the individual is not English and they originate from a culturally different background to the Western culture where the tests were developed (Johnson, Owen, Blackburn, Rehman & Nazroo, 2000). Difficulties occur to varying degrees; some may possess conversational skills in English but may not be clear on the use of psychological language and constructs, complex test instructions or may not be test-wise to western-style tests.

Neuropsychological assessment requires the use of both objective and reliable tests that are designed and adapted in order to adequately measure the cognitive ability of the populations being assessed. Additionally normative data needs to be constructed in order to accurately make clinical judgements about the results of the tests (Bauer, Tobias, & Valenstein, 1993; Mayes, 1986; Squire & Shimamura, 1996). Cross-cultural research has been a neglected area and as such there is a lack of normative published data for people from non-English speaking backgrounds (Ardila, 1995; Fortuny & Mullaney, 1997).

Attempts to make tests more “culturally fair” have been completed in the USA with older African Americans using a range of tests, finding that the measures retained construct validity when adapted for this group (Pedraza et al, 2005). Other tests have been translated into the language of specific groups, with normative data collected, particularly for Spanish speaking groups of Central and South American origin (Goodglass & Kaplan, 1986; Rey, Feldman & Rivas-Vazquez, 1999; Kohnert, Hemandez & Bates, 1998). However, this is not the case for the majority of neuropsychological tests nor the majority of countries using them. Additional difficulties arise when taking the position of recognising culture-specific differences rather than a universalist position of making all tests “culture free” (Nell, 2000) as this requires knowledge of all culturally-specific differences and how these could impact upon certain aspects of different neuropsychological tests.

1.2 Language

Many neuropsychological tests are developed in English and have normative data for English speaking participants. This is despite only approximately 350 million people in the world’s population (of approximately 7 billion) being native English speakers (Kitao, 1999). Even tests not specifically assessing language can be dependent on language knowledge due to the test stimulus, for example Trail Making Test (TMT; Lezak, 1995) which relies on knowledge of the English alphabet. Additionally, the majority of tests have instructions and directions in English, sometimes requiring knowledge of technical terms and complex syntax.

Verbal communication is a requirement for many neuropsychological tests. Alongside the language used in the test manual, an assessor needs to be able to adequately assess the individual’s mood and cooperation (Fortuney & Mullaney, 1998). An understanding of language in terms of these constructs is therefore necessary to avoid misunderstandings particularly in relation to risk and consent.

Bilingual individuals with English as a second language may be assumed to be proficient enough in English to complete tests. However, there may not be equal proficiency in both languages and consequently may underperform if not assessed in their dominant language (Manuel-Dupont, Ardila, Rosselli & Puente, 1992). Additionally, being bilingual in itself could impact on normal performance expectations. On a variety of language-focused measures, English dominant bilinguals appear to be disadvantaged compared to monolinguals regardless of which language they are tested in. (Rosselli, et al. 2000).

1.3 Culture

Environmentalist theories state that the culture and social communities in which one lives can determine and shape the way in which a person develops and conceptualises ideas in the mind (Vygotsky, 1978). Differences in culture and social history were reported by Vygotsky as vitally important to the study of neuropsychology; identifying the impact of culture on the development and functioning of the brain. Neurological investigations indicate that different areas of the brain, used for differing functions, may be dependent on the particular cultural context in which the brain develops (Vygotsky, 1960). This has implications for neuropsychological tests which identify strengths and deficits in cognitive functioning. Diagnoses are often made dependent on the areas of functioning and their related brain regions which are identified in the culture where the test was developed.

Luria (1979) conducted research with different groups of people in Uzbekistan. He identified that people within one culture who had different social experiences, and gave them the same tasks to complete. The research demonstrated that the tasks were completed differently dependent on a person’s educational and social background. Classification and discrimination of objects varied dependent on the knowledge and value set of the individuals assessed. Through qualitative exploration of the responses given, an understanding of the cognitive processes underlying them was established. For those with no formal education and who were not engaged in modern social activity, the classification of objects only made sense if it related to their practical experience. For those with some formal education, the use of abstract thinking during the categorisation of objects was more congruent with what would be considered “average” or "normal" in traditional neuropsychological terminology. Gilbert, (1986) aimed to replicate Luria's work in a region of South Africa. He used similar groupings of people and found that the least educated grouped objects in a way consistent with Luria's findings. On the categorisation task there were mixed findings with some in the educated groups being able to demonstrate abstract thinking when categorising objects. However in a number of cases this required prompting to move from concrete to abstract thought. These studies suggest that even within one culture, social experience can determine performance on measures of cognitive functioning.

Taken alone, without qualitative exploration of responses or knowledge of cultural or social background, scores on neuropsychological tests could be considered meaningless. Additionally, information processing theories such as the triarchic theory of intelligence (Sternberg, 1984) suggest that intelligence relates to the context (culture) in which it is based. Therefore a test measuring a construct related to the deduction of rules used in a context where deduction is used routinely would provide a measurable construct. In cultures where deductive skills are not applied in everyday life, the results of the test would have little or no meaning.

Nativist theory holds that only 20-40% of individual intelligence is influenced by culture; the rest shaped by genetically determined factors (Nell, 2000). Nativists view differences in neuropsychological test scores as a product of these genetic factors, rather than being related to cultural influences (Herrnstein & Murray, 1994). This is a view that has been contentiously argued (Nell, 2000).

Typically neuropsychological tests are created in Western cultures; traditionally described as individualistic societies (Triandis, 1995) which value individual achievement (McClelland, 1961). These concepts alongside the Western educational system lead to the development of test-wiseness (Nell, 2000), that is, the skills and practices learned in childhood educational settings necessary for taking a test. The number of years in education has been demonstrated to be a predictor of performance on the Symmetry Completion test (Crawford-Nutt, 1977a, 1977b, 1977c); more years in education was predictive of a higher score with individuals from the same cultural and ethnic background. Research looking at young children has suggested that prior to the age at which formal schooling begins, differences on intelligence scales between children from differing cultural backgrounds is minimal (Verster & Prinsloo, 1988; Richter-Strydom & Griesel, 1984). Cultural factors including formal education are believed to make the biggest contribution to variance in performance on neuropsychological tests, and these are likely to exist within cultures, as well as between them (Nell, 2000). This suggests that variance will occur between communities and cultures within Western societies where the majority of tests are developed. Test-wise individuals are likely to use automated processes on the majority of neuropsychological tests with fewer degrees of novelty in taking the test. Those subjects who are not familiar with tests will have higher degrees of novelty in both the situation and the test constructs themselves. The same test therefore is likely to be measuring different skills in the two individuals (Sternberg, 1984).

Many neuropsychological tests are literally translated into other languages without regard for cultural relevance. If the test is developed in a Western country and directly translated into a non-Western language, it is unlikely that the constructs underlying the test are shared by the test-maker and test-taker (Nell, 2000). Certain concepts such as asking for the season may be inappropriate in certain cultural contexts, for example those which have only rainy or dry seasons or those where differences in seasons are unnoticeable rendering the notion of seasons irrelevant. Use of inappropriate measures which do not take into account cultural differences can lead to diagnostic errors. In order to avoid this it has been proposed that test items should be correctly developed or adapted with normative data obtained for each population being assessed (Ostrosky-Solis et al.2007).

1.4 Rationale for this review

Due to the complexities in understanding and interpreting the impact of language and culture in the field of neuropsychological testing, the following review aims to provide a critical overview of research which has explored difference as a result of language and culture on a variety of neuropsychological tests.

2. Methodology

A systematic literature search using the search terms: (Neuropsych* Assess* OR Psychometrics OR Neuropsych* test*) AND (culture fair* OR culture fair intelligence test OR cultural test bias) AND (language OR English as a second language) was completed on the databases Psycinfo, Medline and Web of Science. Abstracts and references of identified papers were checked. The search yielded 113 results.

Papers were included if: a) the main aim of the study related to neuropsychological testing; b) the study was based on healthy working-age adults; c) the neuropsychological test was based on English speaking norms; d) the study made reference to non-English culture, and e) the study was published in English.

Applying the above inclusion criteria resulted in 14 papers being selected for review (Figure 1). The key features of each paper are provided in Appendix 1. A description of each paper and a critical evaluation based on guidelines by Greenhalgh (1997) and von Elm, Altman et al. (2008) follows in the main text of the review.

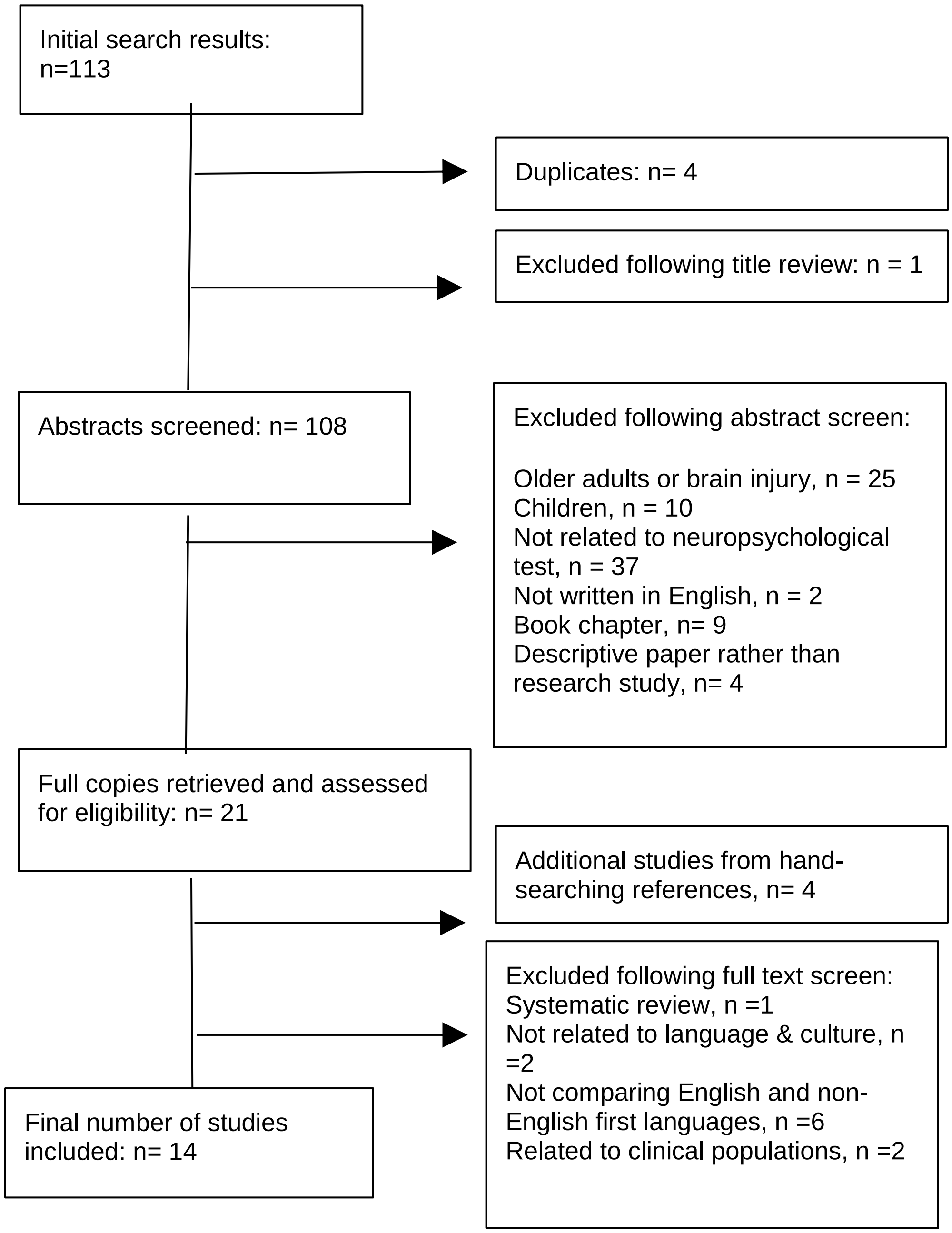


Figure 1: Flow chart of methodological process for selecting papers to review

3. Literature Review

For each of the reviewed studies demographic details and the neuropsychological tests used are provided in Appendix 1.

3.1 Translated versions of neuropsychological tests

Four studies looked at the impact of translating a neuropsychological test into another language when compared to the original normative studies or participants with English as a first language. One of the studies used a standardised translated test; Spanish version of the Wechsler Adult Intelligence Scale, 3rd Edition (WAIS-III) [Renteria, Li & Pliskin, 2008]. Three of the studies translated tests for the purpose of the research project; the Wechsler Memory Scale - Revised (WMS-R) for Spanish speaking people (Demsky, Mittenberg, Quintar, Katell & Golden, 1998), Luria’s Neuropsychological Inventory (LNI) for Zulu-speaking people (Tollman & Msengana, 1990) and the Trail Making Test (TMT) and Colour Trails Test (CTT1 & CTT2) for Chinese speaking people (Lee, Cheung, Chan & Chan, 2000).

Renteria et al., (2008) aimed to explore the reliability and validity of the Spanish WAIS-III when used with Spanish-speakers in the U.S.A. It was found to have adequate reliability on all subtests apart from Letter-Number Sequencing. They hypothesise that this may be due to the letters ‘b’ and ‘v’ being pronounced differently by European Spanish and Latin-American Spanish-speakers. As a result of this finding, the authors recommend that this subtest be eliminated or scored more leniently when testing Latin-American Spanish-speakers. WAIS-III performance was compared to that of other tests given, which were associated with attention and concentration (CTT1 & CTT2). Given that the WAIS-III is not designed to measure these specific abilities; no strong correlations were anticipated. However, the z-test analyses revealed a number of strong correlations with WAIS-III indices, suggesting the translated WAIS-III may tap into attentional constructs, raising concerns regarding the construct validity.

Although there was evidence of some criterion and convergent validity, further examination using paired-sample t-tests found that the Spanish WAIS-III significantly underestimated performance when compared with the Standard Progressive Matrices, Rey Figure Copy, Controlled Oral Word Association Test (COWA) and CTT1 & CTT2. Overall the authors conclude evidence for reliability and validity in some areas for the Spanish version of the WAIS-III. They caution use of certain subtests with Latin-Spanish speakers particularly given that overall cognitive ability could be underestimated.

Demsky et al. (1998) investigated the impact of using a Spanish translated version of the WMS-R with Spanish-speakers, applying published WMS-R norms developed using English-speaking participants. The translated version of the test was adapted to ensure cultural relevance, for example the use of Spanish rather than English names.

The results indicated that the Spanish-speakers scored significantly lower than English-speakers on 10-out-of-13 subtests. When converting the subtest scores to indices, again the Spanish-speakers scored significantly lower. Spanish and English speakers did not differ significantly in terms of age or occupation. However, English-speakers had a significantly higher number of years in education. Statistical corrections indicated that this difference did not impact on the discrepancy between the groups on the WMS-R scores, suggesting that the variances cannot be attributed to education alone.

The results of this study indicate that direct translation of a test, even with consideration for alterations based on language and culture, can result in lower than expected results when compared to a normative population. Lower scores were seen on both verbal and non-verbal subtests indicating that non-verbal subtests of the WMS-R cannot be assumed to be culture free.

Tollman and Msengana (1990) aimed to assess the feasibility of using Luria's Neuropsychological Investigation (LNI) with Zulu-speaking people. The test was translated by a bilingual Zulu and English-speaking individual. The results of the translation of the test indicated a number of adaptations were necessary to validate the test for use with Zulu-speaking people in terms of language, culture and education. There was limited information in the paper concerning the results of these investigations; the main finding reported was that an increased number of years in education were an influencing factor on performance. A detailed account of language and cultural considerations for translating a test into Zulu was given. This offered recommendations in relation to the assessor's knowledge of the culture, in order to help make accurate judgements about the individual's behavioural profile. Though useful at a clinical level, as a research paper there was very limited information about the analysis, results and implications of this research.

Lee et al. (2000) aimed to explore the impact of a Chinese or English first language on the TMT (A & B) and CTT (1 & 2). The test instructions were translated into Chinese for Chinese-English bilingual (CEB) participants. Analysis of the impact of gender, age and education found no significant effects. Significant differences were seen between the CEB and English monolingual (EM) groups on the TMT-A only, with the CEB group performing significantly worse. Further analysis identified significant differences in the time taken to complete the TMT and CTT, with the CTT taking longer for both the CEB and EM groups. The authors concluded that both the TMT and CTT are "fair" tests to use with Chinese-English bilinguals as there was equivalent performance on the more difficult TMT-B. The authors commented on the surprising finding that a difference favouring English monolinguals was found on the simpler component of the TMT-A rather than the more complex TMT-B. They suggested that further research is needed to establish whether the layout of TMT-A favours English monolingual participants.

Summary

Results of the four studies looking at the impact of translated tests suggest that when primarily non-verbal tests are used in cultures where the levels of education are similar, differences on the tests are not significant. In cultures where educational systems are different (e.g. Tollman & Msengana, 1990) more adaptations to the original structure may be necessary which could impact upon the validity of the test. Evidence from Renteria et al. (2008) suggests that translation of a test may impact on the construct validity of the test. Demsky et al. (1998) found that Spanish-speaker’s results were below the English normative sample on the translated WMS-R even when cultural adaptations were made. These studies suggest that careful consideration of language, as well as culture, is necessary when translating tests into another language.

3.2 Cultural familiarity with test items

Three studies investigated the impact of cultural familiarity with the items on the tests used. The tests studied included: Shopping List Memory Tests (SLMTs) in cultures outside of the USA (Lim et al. 2009); the Boston Naming Test (BNT) and Cookie Theft Test (CTTest) for the South African culture (Mosdell, Balchin & Ameen, 2010) and the WAIS-III for the South African culture (Cockcroft, Alloway, Copello & Milligan, 2015).

Lim et al. (2009) aimed to investigate the use of SLMTs across cultures outside the USA. Verbal memory performance was compared across cultural groups using shopping lists identified as culturally appropriate for the population being tested. The use of shopping lists was deemed cross-culturally appropriate as the acquisition of food is common across cultures.

The experiments compared performance on SLMTs across a variety of conditions, including: a) across Western cultures (American and Australian) with the same language; b) across Western cultures with different languages (French and Australian); c) using a directly translated list (Australian to French); d) across non-Western cultures; e) across non-Western cultures with bilingual speakers. Word familiarity appeared to be the most important factor when comparing performance across western cultures which used the same language. However, cultural familiarity appeared to become more important for both Western and non-Western cultures if the same language was not spoken.

Overall the results of the experiments indicated that familiarity with the test items given in the individual’s native language is among the most important factors to consider when conducting a SLMT. Lists should be specific to the language and culture of the individual being tested, with the results of the five experiments suggesting that the specific words used in the lists had less of an impact on performance compared to their cultural familiarity.

Mosdell et al. (2010) aimed to develop culturally relevant versions of the Boston Naming Test (BNT: Kaplan, Goodglass & Weintraub, 1983) and the Cookie Theft Test (CTTest: Goodglass & Kaplan, 1983) for South African populations.

The BNT and CTTest were subject to review by cultural and language experts, as well as members of the community. This was to help create items that were linguistically and semantically similar, but with increased cultural familiarity to the populations being tested. Three groups of participants (English, Afrikaans & isiXhosa speakers) were tested using the culturally adapted versions of the tests named the Groote Schuur Naming Test (GSNT) and the Washing Line Picture Test (WLPT) as well as the BNT and CTTest.

The results indicated that there were significant differences on performances of the BNT and GSNT with larger differences (favouring English-speakers) on the BNT. All participants scored significantly higher on the GSNT and mean scores were more similar across the three groups and in line with BNT norms in the USA. Qualitative analysis revealed that some of the items on the GSNT were unfamiliar to certain groups; however significantly less than on the BNT. Participants reported higher levels of familiarity with items on the GSNT compared to the BNT.

On the WLPT participants reported higher levels of familiarity with the adapted test, though performance on both this test and the CTTest was comparable. The authors comment that an advantage of the CTTest is interpreting narrative-speech ability that is familiar to the person being tested (Mosdell, Balchin & Ameen, 2010; Goodglass & Kaplan, 1983). The WLPT therefore with its increased familiarity is reported as an appropriate test to use with this population.

The authors concluded that this pilot study of the GSNT and WLPT provides evidence to suggest that the two tests are more culturally appropriate for patients with a South African cultural background.

Cockcroft et al. (2015) aimed to compare WAIS-III performance between British university students whose first language was English and black South African university students for whom English was their second language. All South African participants were multilingual, speaking an average of four languages; an African language as their first language and for the majority of the sample English as their second language. For all, English language was acquired in childhood. The entire South African sample was reported to be from low socio-economic backgrounds.

All participants were administered the WAIS-III; the South African sample were administered the WAIS-III according to the South African item administration (Claassen, Krynauw, Paterson, & Mathe, 2001). Exploration of potential cultural differences between the two groups found that British students scored significantly higher than the South African students on the Verbal Comprehension (VCI) and Perceptual Organisation (POI) indices. The South African students scored significantly higher than the British students on the Processing Speed Index (PSI). There were no significant group differences on the Working Memory Index (WMI).

The authors report that there was an impact of low socio-cultural factors including English being a second language and degree of test-wiseness. The Matrix Reasoning subtest particularly, as well as most WMI subtests appeared to show the least amount of cultural difference. The results additionally indicated that the majority of subtests on the WAIS-III were subject to cultural bias, even when the test was adapted for the given culture. This was not limited to purely verbal subtests; non-verbal subtests such as Picture Completion and Block Design also showed cultural bias. The authors discuss this in relation to previous evidence suggesting that nonverbal tests are not by nature, necessarily more culturally fair than verbal ones (Rosselli & Ardila, 2003; Shuttleworth-Edwards, et al., 2013).

Summary

The above studies suggest that cultural familiarity with test items is an important factor in determining performance on the studied neuropsychological tests. As with the previous section, issues arise when tests are simply translated (Lim, et al. 2009) without regard for cultural relevance. There is some evidence to suggest that current tests of working memory may be more culturally relevant than other aspects of the overall test, however more research is needed in this area to establish if this is the case.

3.3 Cultural factors related to the impact of level and quality of education

Two studies explored culturally-related factors in relation to the level and quality of education. Shuttleworth-Edwards, Donnelly, Reid, and Radloff (2004a) explored this in relation to the WAIS-III Digit Symbol Coding subtest, and Shuttleworth-Edwards, Kemp, Rust, Muirhead, Hartman, and Radloff (2004b) in relation to the whole WAIS-III. Both studies looked at differences between those who spoke English as a first language and those who spoke a Black African first language. Level ("grade 12" - 12 years in education and "graduate" - 15+ years in education) and quality of education (advantaged or disadvantaged) were compared within the groups. All participants were administered the English Digit Symbol Coding task or WAIS-III with cultural amendments made where necessary (Claassen et al. 2001).

The results of the Digit Symbol Coding study indicated no effect or interaction of language, level or quality of education on the pairing or free recall aspects of the tests. The authors conclude that this provides evidence that the Digit Symbol Coding task is not significantly impacted upon by language, or education suggesting a level of cultural fairness for this task. Additionally, normative data for the age-group studied were created and found not to differ significantly from the WAIS-III normative data for this age. This test could therefore be assumed to be valid for African populations with advantaged or disadvantaged educational backgrounds who have 12 or more years in education.

The results of the whole WAIS-III study were compared to the WAIS-III English-speaking norms. The results showed a significant effect for level and quality of education. Poorer performances were shown by those in the low level education groups for both English and black African first-language speakers, and those black African first-language speakers in the disadvantaged education group. Quality of education (advantaged/disadvantaged) in the black African first-language speaking group had more of an impact than level of education. For the black African first-language speakers, both performance IQ and verbal IQ were impacted on by level and quality of education. For the English first-language speakers effects of low level education impacted on verbal IQ only.

The authors concluded that low level of education impacts on scores on the WAIS-III, and also that quality of education is an equally, if not more important factor. Additionally the study demonstrated that test-wiseness was a key factor influencing performance on psychological tests and this may be more important than language or crystallised intelligence. This was demonstrated through the equivalently low scores on performance tasks, which were nearly as low as the verbal tasks for black African first-language speakers with low level and quality education.

Summary

The results of these two studies suggest that taken alone, the Digit Symbol Coding task from the WAIS-III is not impacted upon by level or quality of education, however when the whole WAIS-III is considered, significant effects determined by level and quality of education are seen; with lower levels producing lower scores even on tasks not reliant on crystallised intelligence.

3.4 Completing test in a second language

Five studies looked at the impact of completing a test in a person’s second language. Dugbartey, Townes, and Mahurin (2000), looked at the TMT and CTT with people who spoke

Turkish as a first language. Lu and Bigler (2000) also looked at the TMT with people whose first language was Chinese. Harris, Cullum and Puente (1995) looked at differences on a verbal-learning-memory test with people speaking Spanish as a first language. Carstairs, Myors, Shores and Fogarty (2006) used a neuropsychological test battery with people from a non-English-speaking background and Razani, Burciaga, Madore and Wong (2007) used a neuropsychological test battery with people from ethnically-diverse backgrounds. All were compared to people speaking English as a first language.

Dugbartey et al. (2000) administered a neuropsychological-test battery to participants speaking English or Turkish as a first language. Analysis indicated a lack of significant differences between TMT-A and CTT-1, and performance on these tasks was significantly correlated. There was however, a significant difference between TMT-B and CTT-2 and these tests were reported to be only moderately correlated.

The authors concluded that this study provides evidence to suggest that the CTT-2 is not a culturally equivalent version of the TMT-B, possibly tapping into different cognitive skills. They suggest that this may relate to the increased number of stimuli on the CTT-2 which adds additional scanning and visually distracting material as well as cognitive flexibility; increasing requirements for quick performance on the task.

Lu and Bigler (2000) aimed to establish whether there was a bias on the TMT with a native language which does not use the English alphabet. All participants were administered the Digit Symbol Coding task (WAIS-III) and the TMT-A and TMT-B. The Chinese participants were also tested with the TMT-B using Chinese characters.

Independent sample t-tests revealed that there were significant differences between the groups on the TMT-B when using the original measure. However, there were no differences when comparing the English speaking participants tested with the English TMT-B and Chinese participants tested with the Chinese adaptation of the TMT-B. There were no significant differences using TMT-A or the Digit Symbol Coding Task.

The results of the study suggest that the Digit Symbol Coding Task and TMT-A can be used with Chinese people who speak English as a second language. The TMT Part-B when used with Chinese who speak English as a second language could lead to misleading results therefore the Chinese adaptation of TMT-B could be a linguistically-equivalent test with this population.

Harris et al. (1995) aimed to establish the impact of bilingualism on a verbal learning memory test with “balanced” (equal proficiency in both languages, N=22) and “non-balanced” (dominant in one language, N=22) bilinguals, compared to English monolinguals. Additionally, they aimed to develop English and Spanish equivalent verbal learning lists.

All participants were administered the Raven’s Coloured Progressive Matrices (CPM: Raven, 1956) to screen for intellectual impairment. This test is reported to have reliability across ethnic groups (Birkmeyer, 1965; Jensen, 1974; Valencia, 1979, 1984; Carlson & Jensen, 1981). Two verbal learning lists were constructed in English and Spanish. Measures were taken to ensure that the construction of the lists was in accordance with a validated measure, the California Verbal Learning Test (CVLT: Delis, Kramer, Kaplan & Ober, 1987).

Results of the screening test indicated that the non-balanced individuals performed significantly worse on the CPM than balanced individuals. On the list-learning test, non-balanced individuals clustered words significantly less than balanced individuals, but to a similar level as monolinguals. Non-balanced individuals also recalled significantly fewer words on the English list than balanced individuals and monolinguals. When tested in their dominant languages the three groups did not differ in terms of clustering or word recall.

The authors concluded that degree of bilingualism is an important variable in relation to learning a verbal word list. Those individuals tested in their non-dominant language recalled significantly fewer words despite similar levels of clustering as monolinguals. As non-balanced individuals did not differ from the other groups when tested in their dominant language, there does not appear to be a difference in memory, rather a disadvantage when tested in their non-dominant but proficiently spoken second language. The authors suggested that the use of verbal learning lists across cultures has equivalence as long as the individual is tested with a list of culturally relevant words in their dominant language. The difference in CPM scores was considered in relation to the quality of education that may have been received by the non-balanced bilingual individuals, who received their education in Mexico, an area with wide variability in educational services. This is discussed in relation to familiarity with Western test-taking procedures which could have impacted upon the non-balanced individuals performances.

Carstairs et al. (2006) compared participants from an English speaking background (ESB) with participants of non-English speaking backgrounds (NESB). The participants of non-English speaking origins were categorised further into those who first spoke a language other than English as a child (NESB-OE) and those who first spoke English as a child (NESB-E). There were no significant differences in age, gender, and years in education or level of qualification between the three groups.

All participants were administered a neuropsychological test battery. ANOVA's were computed for the data and the results indicated that participants from NESB-OE had significantly lower scores on the WAIS-R (VIQ, PIQ and FSIQ) than both other groups. Additionally, the NESB-E group had significantly lower scores on the WAIS-R (PIQ) than those from ESB. The authors suggested that language differences may have contributed to the differences with VIQ, and cultural differences for the PIQ. Further evidence for the impact of language proficiency was indicated through performance on the SCOLP (Speed and Capacity of Language Processing Test, Form-A) which assessed speed of (English) language processing. Those from NESB-OE had significantly lower scores than the other two groups and the authors suggested this reflected a disadvantage resulting from lower levels of language proficiency in English. The authors concluded that individuals from a NESB with a first language other than English are placed at a disadvantage when tested in English, particularly with verbal tasks.

Razani et al. (2007) aimed to investigate the influence of language and acculturation on neuropsychological test performance (specifically information processing and attention) with a group of monolingual English speaking Anglo-Americans (MEAA) and "ethnically-diverse" (ED) individuals.

The ED group differed in terms of: languages spoken; time spent in the U.S; time spent speaking English and proportion of education inside/outside of the U.S. All ED participants completed a measure of acculturation; the Acculturation Rating Scale for Mexican Americans (ARSMA: Cuellar, Harris, & Jasso, 1980) with terminology adapted dependent on the origin of the individual tested.

Participants were administered a neuropsychological test battery and all were tested in English. The authors noted that due to "various reasons" the entire ED group did not complete all measures. The impact of this is not acknowledged in the results section. Due to differences in age between the two groups, age was used as a covariate in all of the analyses. These revealed that the MEAA group performed better than the ED group on Digit Span, Digit Symbol, TMT- B, Stroop B, Stroop C, and Auditory Consonant Trigrams (ACT) 18-s delay. There were no differences for the less complex TMT, Stroop or ACT tasks.

Further analysis exploring whether levels of acculturation had an impact on the tests found that the MEAA group performed better than both ED groups on Digit Span, TMT-B, Stroop B , Stroop C, and ACT-18s delay. There were no significant differences between the high and low acculturated groups, suggesting that level of acculturation did not significantly impact on test performance.

The amount of education obtained outside of the U.S correlated with a number of measures, indicating poorer performance on the Digit Symbol, Digit Span, TMT A & B and Stroop B. The authors discussed this in relation to the impact of the quality of education being a predictor of neuropsychological test performance which has previously been demonstrated with older adults (Manly, Byrd, & Pegah, 2004). The authors discussed the overall results of the study in relation to the impact of completing education in countries where test-taking expectations are comparable to those where the test is taking place. Additionally, they discuss the assumptions that are made about individuals who speak fluent English; suggesting that despite good fluency in a second language, performance on measures of information processing and attention (which have relatively low verbal mediation) can still be negatively impacted upon. This has implications for testing individuals with English as a second language, regardless of their level of proficiency in this.

Summary

The results of studies looking at the impact of completing a test in a second language suggest that tests requiring knowledge of English characters may prove more difficult for those with English as a second language, even when the verbal demands for the test are quite low (e.g. CTT and TMT). When the tests are adapted to reflect the individual's dominant linguistic background, differences between groups are reduced (e.g. Lu & Bigler, 2000). Furthermore, even when language proficiency in a second language is good, performances appear worse compared with people taking the tests in their first language. This indicates that factors other than language proficiency may impact on performance when taking tests in second languages. These factors may include education and level of test-wiseness as discussed by some of the studies (e.g. Harris et al., 1995; Razani et al., 2007).

3.5 Methodological critique

Sampling: Overall the studies had good descriptions of the samples as well as inclusion and exclusion criteria for participation. Screening for neurological factors that may influence cognitive performance (e.g. head injury, stroke) was not discussed in seven-out-of-the-fourteen studies reviewed. This could impact on the validity of the results if such screening was not done.

There were small sample sizes in six of the studies which could lead to type II errors affecting the validity of the results. Additionally, this impacts upon the generalisability of the findings.

Generally, sample sizes were not found to reflect the lifespan. Overall, use of wider age ranges in the studies would increase the ability to generalise the findings to the wider population. Four studies use primarily students, again this has implications for the generalisability of the findings as well as limiting it to those who are well educated and practiced in test-taking.

The majority of the reviewed studies use participants with high levels of education which limits the conclusions that can be drawn regarding those with lower levels of education. The implications of this are discussed in a number of the studies particularly those based in South Africa where differences in education vary vastly. The only studies with wider ranges of years in education are Renteria et al. (2008), Harris et al. (1995) and Razani et al. (2007). Although Lim et al. (2009) used primarily highly-educated students, they reported that the results were unlikely to be impacted upon if using people from less-educated backgrounds due to the knowledge base requirements of the test used (shopping list memory test).

Cohort controlling: Six-out-of-the-fourteen studies reviewed did not discuss the matching of groups. This has implications for the validity of the results and it is unclear in these cases if the results can be attributed to language or cultural factors or whether other factors such as age, education or gender had an effect. Where matching of groups was discussed, the process of this was described and any significant differences acknowledged.

Statistical analyses: Generally, the analysis used in the studies was well described with details of the initial analysis and any post-hoc analysis used. However, in two of the studies (Tollman & Msenga, 1990; Lu & Bigler, 2000) specific details of the analysis were not provided. This impacts on the reader's ability to make sense of the findings and conclusions, as well as impacting upon the replicability and transparency of the findings.

Three of the studies (Carstairs et al. 2006; Demsky et al. 1998; Dugbartey et al. 2000) report on the use of an alpha rate of 0.01 to reduce the possibility of type I errors. This reduces the possibility the rejection of the null hypothesis is false, increasing the validity of the findings. The alpha level used was not discussed explicitly in the majority of the studies, though was present in some.

Languages and cultures used: In the studies looking at the translation of tests there was no formal test of language ability. This impacts on the ability to generalise the findings. In those studies where participants were residing in an English-speaking country and had been for a number of years, their proficiency in their native language may have been reduced and they may have been at a disadvantage with a translated test. This suggests that choice should be given to the individual, or a language proficiency test created in order to select the most appropriate test.

It also raises questions as to whether bilingual tests should be developed or whether tests should have norms linked to factors important to proficiency, such as: the length of time spent speaking a particular language; use of language; dominance of language, and age the speaker began using their second language.

Renteria et al. (2008) was the only study which discussed test-item bias, and commented on subtests such as Information or Similarities (WAIS-III) which may include knowledge relevant to Spanish culture, but not necessarily to Latin American. Additionally, the order of items from easy to more difficult may not be the same when tested in different cultures; a result found in the Lee et al. (2000), where the “easier” component of the test produced significant differences between the groups, whereas the more complex component did not.

There were differences across studies in the numbers of languages and cultures examined. Though four of the reviewed studies considered a number of different cultures (Harris et al. 1995; Carstairs et al. 2006; Razani et al. 2007; Lim et al. 2009) which is useful for understanding whether cultural background has an impact on the results of the test, it is difficult to make recommendations for using the tests with certain cultural groups due to the heterogeneity of the samples. A number of studies (Lee et al. 2000; Lu & Bigler, 2000; Tollman & Msenga, 1990; Dugbartey et al. 2000; Renteria et al. 2008) used only one language group compared to English-language speaking and whilst this impacts on the generalisability to other language groups, does allow for specific recommendations for using the studied test.

Neuropsychological tests: Generally a good description of the tests used was provided throughout the studies. Only one (Rentira et al. 2008) described the psychometric properties of the tests which again adds to the replicability and validity of the results. For those studies completed in South Africa consideration of the South African standardisation procedures was acknowledged which ensured culturally-appropriate stimuli were used. In those studies where translation of a test was completed, there is information provided relating to the process and considerations taken to ensure both the equivalence to the original test as well as cultural relevance.

4. Discussion

This review aimed to critically appraise the literature exploring the impact of language and culture on neuropsychological tests in healthy working-age adults. Fourteen papers were reviewed across a variety of cultures with different languages and neuropsychological tests used. All of the reviewed studies have conclusions with implications for future research as well as clinical applications of the findings, though in some cases methodological weaknesses impact upon the extent to which the findings could be applied more widely.

The results of the reviewed studies suggest that in the case of some tests, the impact of language and culture is minimal, for example the Trail Making Test. Effects are reduced further when the test is adapted for the specific group being tested. This is an important finding as this test is reported to be commonly used in both forensic (Lees-Haley, Smith, Williams & Dunn, 1996) and clinical (Butler, Retzlaff & Vanderploeg, 1991) settings, as well as in distinguishing those with brain injury from healthy controls (Berg, Franzen, & Wedding, 1987; Eson, Yen & Bourke, 1978; Reitan, 1958; 1971).

Cultural familiarity with the test-items is a key finding of the literature. When tests are adapted to be more in line with the individual's cultural background, differences between groups are less likely. Only the Renteria et al. (2008) addressed limitations in relation to test-item bias and whether the test was measuring the intended constructs. This could significantly impact on the validity of the tests (Sternberg, 1984) and implies that knowledge of how familiar constructs are within cultures is necessary for meaningful neuropsychological interpretation.

Previous research has suggested that non-verbal tests are less likely to be subject to language or cultural differences. Demsky et al. (1998) found evidence in their study to suggest that non-verbal tests were not culture free. Additionally, Cockcroft, et al. (2015) found that non-verbal subtests on the WAIS-III were subject to the influence of language and culture and that tests of working memory were the least likely subtests to be influenced by different cultural backgrounds. This is particularly pertinent given that such tests measure fluid intelligence, that is not explicitly taught or knowledge driven and therefore should be less influenced by education.

When tests are translated into another language, differences between groups appear to be minimal if cultural factors such as the quality of education are controlled for. In cases where the cultural backgrounds of the test and the language to be translated to are very different, particularly in relation to educational systems, translation can be unhelpful. Furthermore, when a number of adaptations are made to a test during translation, the validity of the test can be impacted upon. For the shopping list memory test, the simplicity of the test allows for direct translation with culturally-familiar items. Such tests appear appropriate to use across a variety of languages and cultures due to the commonality of food acquisition. This links to previous research which suggests that social experience in a culture can impact upon test performance (Luria, 1979; Gilbert, 1984).

A number of studies looked at test performance in countries where there are differences in language and culture (particularly related to education). The level and quality of education was seen to have an impact on test performance particularly in countries such as South Africa where in the past education was segregated according to race. This effect of education has been shown previously with older African-American adults, particularly in relation to the quality of education received (Manly, Jacobs, Touradji, Small & Stern, 2002). Educational differences have additionally been highlighted as one of the biggest influencing factors both within and between cultures (Nell, 2000). These findings highlight the importance of establishing educational history when conducting neuropsychological testing, as well as assessing for premorbid intelligence and degree of test-wiseness; as these were demonstrated by Shuttleworth-Edwards et al. (2004a) as being more important than crystallised intelligence.

The results of the reviewed studies suggest caution against making ability-related assumptions when assessing a person who is perceived to be bilingual. Even those bilinguals who reported equal proficiency in both languages showed reduced performance on tests when compared to monolinguals. There has been some evidence for normative data and guidance produced for Spanish bilingual speakers (for example Ardila, Rosselli, & Puente, 1994) however this is still an area relatively under researched and the impacts of bilingualism on test performance are not fully understood. This guidance and many of the studies are completed in the U.S.A. with Spanish-speakers. However, differences in Spanish-speaking backgrounds are not always acknowledged. Although Spanish-speakers may be living in the U.S.A. they may have received their education in a country with differing quality of schooling. Additionally, within the Spanish language there are differences for example between European-Spanish and Latin-Spanish. Without this knowledge individuals may be scored negatively and results interpreted as impairment of cognitive functioning rather than as a result of language.

This review explored issues relating to language and culture within a working-age population and therefore limits the extent to which the impacts of language and culture can be extended to other groups such as children, older adults and clinical populations which, if explored could add value to the understanding of the impact of language and culture in relation to neuropsychological tests in different clinical areas. Additionally, the meaning of "culture" in relation to identifying papers with culture in their content is defined for the purposes of this review by both the author of this review as well as the authors of the reviewed papers. Thus meanings of culture may be different dependent on how the person conceptualises culture as meaningful for them. It is difficult to know therefore if identifying culture encapsulates the same meaning from person to person especially when considering factors such as the differences between individualistic and collectivist cultures.

5. Future Research

Given the importance of developing both linguistically- and culturally-relevant tests to be used cross-culturally there is a need for more research looking at the cultural familiarity of test-items as well as the commonality of word use across languages, particularly the understanding of certain words when spoken in a second language. This research would be necessary for individual tests with recommendations for separate languages and cultures.

Additionally, given the reported importance of education, particularly the quality of this, there is a need for the exploration of the impact of education on neuropsychological tests in the UK. There appears to be a lack of cross-cultural research that has been completed in the UK, and given the upward trend of increasing diversity from migration, it is important to establish the relevance of the neuropsychological tests used with people of languages and cultures originating outside of the UK.

References

Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. Journal of Clinical and Experimental Neuropsychology, 17, 143-150.

Ardila, A., Ostrosky-Solis, F., Rosselli, M., & Gómez, C. (2000). Age-related cognitive decline during normal aging: The complex effect of education. Archives of Clinical Neuropsychology, 15(6), 495-513.

Army Individual Test Battery (1994). Manual of directions and scoring. Washington, DC: War department, Adjutant General's Office.

Baddeley, A., Emslie, H., & Smith, I. N. (1992). The Speed and Capacity of Language Processing test manual. Bury St Edmunds, UK: Thames Valley Test Company.

Bauer, R.M., Tobias, B., & Valenstein, E. (1993). Amnesic disorders. In K.M. Heilman & E. Valenstein (Eds.), Clinical neuropsychology (3rd Ed). New York: Oxford University Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychology Bulletin, 107, 238–246. doi: 10.1007/BF02294170

Benton, A.L. (1969). Development of a multilingual aphasia battery: Progress and problems. Journal of the Neurological Sciences, 9, 39-48.

Benton, A. L., & Hamsher, K. de S. (1976). Multilingual Aphasia Examination. Iowa City, IA: Department of Neurology, University of Iowa Hospitals. [Revised edition, 1978.]

Berg, R. A., Franzen, M. D., & Wedding, D. (1987). Screening for brain impairment. New York: Springer.

Birkmeyer, F. (1965). The relationship between the Coloured Progressive Matrices and the Wechsler Intelligence Scale for Children. Psychology in the Schools, 2, 278-280.

Bisin, A., & Verdier, T. (2000). "Beyond the melting pot": Cultural transmission, marriage, and the evolution of ethnic and religious traits. Quarterly Journal of Economics, 115(3), 955-988. doi: 10.1162/003355300554953

Butler, M., Retzlaff, P., & Vanderploeg, R. (1991). Neuropsychological test usage. Professional Psychology: Research and Practice, 22, 510–512.

Cambridge Cognition. (1995). Cambridge Neuropsychological Test Automated Battery (CANTAB) Manual. Cambridge: Author.

Carlson, J.S. & Jensen, C.M. (1981). Reliability of the Raven's Coloured Progressive Matrices test: Age and ethnic group comparisons. Journal of Consulting and Clinical Psychology, 49, 320-322.

Carstairs, J.R., Myors, B., Shores, E.A., & Fogarty, G. (2006). Influence of language background on tests of cognitive abilities: Australian data, Australian Psychologist, 41(1): 48-54.

Claassen, N. C. W., Krynauw, A. H., Paterson, H., and Mathe, M. (2001). A Standardisation of the WAIS-III for English-Speaking South Africans. Pretoria:
Human Sciences Research Council.

*Cockcroft, K., Alloway, T., Copello, E., & Milligan, R. (2015). A cross-cultural comparison between South African and British students on the Wechsler Adult Intelligence Scales Third Edition (WAIS-III). Frontiers in Psychology, 6:297. doi: 10.3389/fpsyg.2015.00297

Coleman, D. (2013). Immigration, Population and Ethnicity: The UK in International Perspective. Migration Observatory briefing, COMPAS, University of Oxford, UK.

Crawford, J.R., Wright, R., & Bate, A. (1995). Verbal, figural and ideational fluency in CHI. International neuropsychological society and Australian society for the study of brain impairment 2nd Pacific Rim conference, Cairns.

Language, Culture and Neuropsychological Tests

Crawford-Nutt, D.H. (1977a). The test of mental ability among black teachers in Bophuthatswana (Contract report C/PERS 257). Johannesburg: National Institute for Personnel Research.

Crawford-Nutt, D.H. (1977b). The effect of educational level of the test scores of people in South Africa. Psychologia Africana, 17, 49-59.

Crawford-Nutt, D.H. (1977c). The SYMCO test as a test of ability. Psychologia Africana, 17, 79-98.

Cuellar, I., Harris, L.C., & Jasso, R. (1980). An acculturation scale for Mexican American normal and clinical populations. Hispanic Journal of Behavioural Sciences, 2(3), 199-217.

D'Elia, L. F. D., Satz, P., Uchiyama, C. L., & White, T. (1994). Colour trails test. Odessa, FL: Psychological Test Resources.

D'Elia, L.F.D., Satz, P., Uchiyama, C.L., & White, T. (1996). Colour Trails Test. Professional Manual. Odessa, FL: Psychological Test Resources.

Delis, D., Kramer, J.H., Kaplan, E., & Ober, B.A. (1987). The California Verbal Learning Test-Research edition. New York: Psychological Corporation.

*Demsky, Y.I., Mittenberg, W., Quintar, B., Katell, A.D., & Golden C.J. (1998). Bias in the use of standard American norms with Spanish translations of the Wechsler Memory Scale-Revised. Test. 5(2), 115-121.

Dolan, C. V., Roorda, W., and Wicherts, J. M. (2004). Two failures of Spearman's hypothesis: the GAT-B in Holland and the JAT in South Africa. Intelligence. 32, 155–173. doi: 10.1016/j.intell.2003.09.001

Dolan, C. V., Colom, R., Abad, F. J., Wicherts, J. M., Hessen, D. J., and van der Sluis, S. (2006). Multi-group covariance and mean structure modelling of the relationship between the WAIS-III common factors and sex and educational attainment in Spain. Intelligence. 34, 193–210. doi: 10.1016/j.intell.2005.09.003

*Dugbartey, A.T., Townes, B.D., Mahurin, R.K. (2000). Equivalence of the Colour Trails Test and Trail Making Test in non-native English-speakers. Archives of Clinical Neuropsychology, 15 (5), 425-431.

Eson, M. E., Yen, J. K., & Bourke, R. S. (1978). Test of recovery from serious head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 41, 1036–1042.

Fortuny, L.A. & Mullaney, H.A. (1997). Neuropsychology with Spanish speakers: language use and proficiency issues for test development, Journal of Clinical and Experimental Neuropsychology, 19, 615-622.

Fortuny, L.A. & Mullaney, H.A. (1998). Assessing patients whose language you do not know: can the absurd be ethical? The Clinical Neuropsychologist, 12 (1), 113-126.

Freeman, R.D. (1980). Visual acuity is better for letters in rows than in columns. Nature, 286, 62-64.

Freidli, L. & Stearn, R. (2015). Positive affect as coercive strategy: conditionality, activation and the role of psychology in UK government workfare programmes. Critical Medical Humanities, 41, 40-47. doi:10.1136/medhum-2014-010622 .

Golden, J.C. (1978). Stroop Colour and Word Test. Chicago, IL: Stoelting.

Goodglass, H., & Kaplan, E.(1986). La evaluación de la afasia y de trastornos relacionados [The test of aphasia and related disorders] (2nd ed.). Madrid, Spain: Editorial Medica Panamericana.

Gilbert, A. J. (1986). Psychology and social change in the third world: A cognitive perspective, Unpublished doctoral dissertation, University of South Africa, Pretoria, South Africa.

Goodglass, H. & Kaplan, E. (1983). Boston Diagnostic Aphasia Examination (BDAE). Lea and Febiger. Psychological Test Resources, Odessa, FL.

Goodglass, H., & Kaplan, E. (1983). The test of aphasia and other neurological disorders. Baltimore, MD: Williams and Wilkins.

Greenhalgh, T. (1997). How to read a paper. Assessing the methodological quality of published papers. BMJ: British Medical Journal, 315:305. doi: http://dx.doi.org/10.1136/bmj.315.7103.305

*Harris, J.G., Cullum, M., & Puente, A.E. (1995). Effects of bilingualism on verbal learning and memory in Hispanic adults, Journal of International Neuropsychological Society, 1, 10-16.

Hennesy, M. (1995). AUSNART: The development of an Australian version of the NART. In J. Fourez, & N. Page (Eds.), Treatment issues and long term outcomes: Proceedings of the 18th annual brain impairment conference. Hobart: Australia.

Herrnstein, R.J. & Murray, C. (1994). The Bell Curve: Intelligence and class structure in American life. New York: Free Press.

Jensen, A.R. (1974). How biased are culture-loaded tests? Genetic Psychology Monographs, 90, 185-244.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). The Boston Naming Test (2nd ed.). Philadelphia: Lea & Febiger.

Kitao, K. (1996). Why do we teach English? The Internet TESL Journal, 2(4). Retrieved from: http://iteslj.org/Articles/Kitao-WhyTeach.html

Kline, R. B. (1998). Principles and practice of structural equation modelling. New York: The Guilford Press.

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(3), 422-440. doi: 10.1006/brln.1998.2001.

*Lee, T.M.C., Cheung, C.C.Y., Chan, J.K.P., & Chan, C.C.H. (2000). Trail Making across languages, Journal of Clinical and Experimental Neuropsychology, 22(6), 772-778.

Lee, C. (2000). Are trail making and colour trails tests of equivalent constructs? Journal of Clinical and Experimental Neuropsychology. 22(4), 529–534.

Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1996). Forensic neuropsychological test usage: An empirical survey. Archives of Clinical Neuropsychology, 11, 45–51.

Lezak, M.D. (1996). Neuropsychological test (3rd ed.). New York: Oxford University Press.

Lezak, M. D. (2012). Neuropsychological test. Oxford: Oxford University Press.

*Lim, Y. Y., Prang, K. H., Cysique, L., Pietrzak, R. H., Snyder, P. J., & Maruff, P. (2009). A method for cross-cultural adaptation of a verbal memory test. Behavior Research Methods, 41(4), 1190-1200.

*Lu, L. & Bigler, E.D. (2000). Performance on original and a Chinese version of Trail Making Test Part B: A normative bilingual sample, Applied Neuropsychology, 7:4, 243-246. DOI: 10.1207/S15324826AN0704_6

Luria, A.R. (1973). The working brain. New York: Basic Books.

Luria, A. R. (1979). The making of mind, Harvard University Press, Cambridge, MA.

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.

Manly, J.J., Byrd, D.A., & Pegah, T. (2004). Acculturation, reading level, and neuropsychological test performance among African American elders, Applied Neuropsychology, 11, 37-46.

Manuel-Dupont, S., Ardila, A., Rosselli, M., & Puente, A.E. (1992). Bilingualism. In A.H. Puente & R.J. McCaffrey (Eds.), Handbook of neuropsychological test: A biopsychosocial perspective, (pp. 193-210). New York: Plenum.

Marin, G., & Marin, B. V. (1991). Research with Hispanic populations. Newbury Park, CA: Sage Publications.

Mayes, A.R. (1986). Learning and memory disorders and their test. Neuropsychologia, 24, 25-39.

McClelland, D.C. (1961). The achieving society. New York: The Free Press.

Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. Psychological Bulletin, 111, 172–175.

*Mosdell, J. Balchin, R. & Ameen, O. (2010). Adaptation of aphasia tests for neurocognitive screening in South Africa. South African Journal of Psychology, 40 (3), 250-261.

Nell, V. (2000) Cross-cultural neuropsychological test: theory and practice. Erlbaum: New Jersey.

Ostrosky-Solis, F., Gomez-Perez, E.M., Matute, E., Rosselli, M., Ardila, A. & Pineda, D. (2007). Neuropsi Attention and Memory: A neuropsychological test battery in Spanish with norms by age and educational Level. Applied Neuropsychology, 14(3), 156-170, doi: 10.1080/09084280701508655

Pedraza, O., Lucas, J. A., Smith, G. E., Willis, F. B., Graff-Radford, N. R., Ferman, T. J., Ivnik, R. J. (2005). Mayo's older African American normative studies: Confirmatory factor analysis of a core battery. Journal of the International Neuropsychological Society, 11(02). doi: 10.1017/S1355617705050204

Psychological Corporation. (1997). WAIS-III WMS-III technical manual. San Antonio, TX: Psychological Corporation.

Raven, J.C. (1956). The Coloured Progressive Matrices. London: H.K. Lewis & Co.

Raven, J. C., Court, J. H., & Raven, J. (1992). Standard progressive matrices manual. Oxford, UK: Oxford Psychologists Press.

*Razani, J., Burciaga, J., Madore, M., & Wong, J. (2007). Effects of acculturation on tests of attention and information processing in an ethnically diverse group. Archives of Clinical Neuropsychology, 22, 333-341.

Reitan, R. M. (1958). The validity of the Trail Making Test as an indicator of organic brain damage. Perceptual and Motor Skills, 9, 127–130.

Reitan, R. M. (1971). Trail Making Test results for normal and brain-damaged children. Perceptual and Motor Skills, 33, 575–581.

*Renteria, L., Li, S. T., & Pliskin, N. H. (2008). Reliability and validity of the Spanish language Wechsler adult intelligence scale in a sample of American, urban, Spanish-speaking Hispanics. The Clinical Neuropsychologist, 22(3), 455-470. DOI: 10.1080/13854040701336428

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rey, G. J., Feldman, E., Rivas-Vazquez, R., Levin, B. E., & Benton, A. (1999). Neuropsychological test development and normative data on Hispanics. Archives of Clinical Neuropsychology, 14(7), 593-601. doi: 10.1093/arclin/14.7.593.

Richter-Strydom, L.M. & Griesel, R.D. (1984). African infant precocity: A study of group of South African infants from two to fifteen months of age. Pretoria, South Africa: University of South Africa.

*Rivera Mindt, M.R., Arentoft, A., Germano, K. K., D'Aquila, E., Scheiner, D., Pizzirusso, M., & Gollan, T. H. (2008). Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. Neuropsychology review, 18(3), 255-268. doi: 10.1007/s11065-008-9069-7.

Language, Culture and Neuropsychological Tests

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solí, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. Applied Neuropsychology, 7(1), 17-24.

Rosselli, M., and Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. Brain Cognition. 52, 326–333. doi: 10.1016/S0278-2626(03)00170-2

*Shuttleworth-Edwards, A.B., Donnelly, M.J., Reid, I., & Radloff, S.E. (2004). A Cross-cultural study with culture fair normative indications on WAIS-III Digit Symbol—Incidental Learning. Journal of Clinical and Experimental Neuropsychology, 26(7), 921-932. doi 1380-3395/04/2607-921

*Shuttleworth-Edwards, A.B., Kemp, R.D., Rust, A.L., Muirhead, J.G.L., Hartman, N.P., & Radloff, S.E. (2004) Cross cultural effects of IQ test performance: A review and preliminary normative indications on WAIS-III test performance. Journal of Clinical and Experimental Neuropsychology. 26: 7 DOI: 10.1080/13803390490510824

Shuttleworth-Edwards, A. B., Gaylard, E. K., and Radloff, S. E. (2013). “WAIS-III test performance in the South African context: extension of a prior cross-cultural normative database,” in Laher, S. & Cockcroft, K. (eds). Psychological test in South Africa: Research and applications. Johannesburg: Wits University Press.

Silver, C. H., Ruff, R. M., Iverson, G. L., Barth, J. T., Broshek, D. K., Bush, S.S., Koffler, S.P. & Reynolds, C.R. (2008). Learning disabilities: The need for neuropsychological evaluation. Archives of Clinical Neuropsychology, 23(2), 217-219 doi: 10.1016/j.acn.2007.09.006.

Silver, C. H., Blackburn, L. B., Arffa, S., Barth, J. T., Bush, S. S., Koffler, S. P., Pliskin, N.H., Reynolds, C.R., Ruff, R.M., Troster, A.I., Moser, R. S. & Elliot, R.W. (2006). The importance of neuropsychological assessment for the evaluation of childhood learning disorders: NAN Policy and Planning Committee. Archives of Clinical Neuropsychology, 21(7), 741-744. doi:10.1016/j.acn.2006.08.006

Snodgrass, J.G., & Vanderwart, M. (1980). A standard set of 260 pictures: norms for naming, agreement, familiarity and visual complexity. Journal of Experimental Psychology: Human Learning and Memory. 6, 174-215.

Squire, L.R., & Shimamura, A. (1996). The neuropsychology of memory dysfunction and its test. In I. Grant & K.M. Adams (Eds.), Neuropsychological test of Neuropsychiatric disorders (pp. 232-262). New York: Oxford University Press.

Sternberg, R. J. (1984). Toward a triarchic theory of human intelligence. Behavioural and Brain Sciences, 7, 269–287.

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

TEA Ediciones. (2001). WAIS-III Escala de Inteligencia de Wechsler para Adultos–III: manual técnico. Madrid, Espana: TEA Ediciones.

*Tollman, S.G., & Msengana, N.B. (1990). Neuropsychological test: Problems in evaluating the higher mental functioning of Zulu-speaking people using traditional western techniques. South African Journal of Psychology, 20 (1).

Triandis, H.C. (1995). Individualism and Collectivism. Boulder: CO: Westview Press.

Valencia, R.R. (1979). Comparison of intellectual performance of Chicano and Anglo third-grade boys on the Raven's Coloured Progressive Matrices. Psychology in the Schools, 16, 448-453.

Valencia, R.R. (1984). Reliability of the Raven's Coloured Progressive Matrices for Anglo and for Mexican-American children, Psychology in the Schools, 21, 49-52.

Language, Culture and Neuropsychological Tests

Verster, J.M. & Prinsloo, R.J. (1988). The diminishing test performance gap between English speakers and Afrikaans speakers in South Africa. In S.H. Irvine & J.W. Berry (Eds), Human abilities in cultural context (pp. 534-560). Cambridge: Cambridge University Press.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P., & STROBE initiative (2008). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Journal of Clinical Epidemiology. 61(4):344-9. doi:10.1016/j.jclinepi.2007.11.008.

Vygotsky, L.S. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1960). Razvitie vysshikh psikhicheskikh funktsii [The development of the higher mental functions]. Moscow: Akad. Ped. Nauk. RSFSR.

Wicherts, J. M., and Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: an illustration using IQ test performance of minorities. Educational Measurement Issues and Practice. 29, 39–47. doi: 10.1111/j.1745-3992.2010.=
00182.x

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

Wechsler, D. (1987). Wechsler Memory Scale - Revised Manual. San Antonio: The Psychological Corporation.

Wechsler, D. (1989). Wechsler Intelligence Scale for Children-III manual. New York: The Psychological Corporation.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale-Third Edition. London: The Psychological Corporation Limited.

Section B

An exploration of the impact of language and culture on a neuropsychological test

Claire Haddlesey

April 2016

Word Count: 7129

A thesis submitted in partial fulfilment of the requirements of Canterbury Christ Church University for the degree of Doctor of Clinical Psychology

For submission to Neuropsychological Rehabilitation

Abstract

Introduction: The use of neuropsychological tests with regards to language and culture is a relatively under researched area. Previous research has found that cultural familiarity with test-items is important for minimising differences between groups and that completing tests in a second language can increase differences between groups on both language- and performance-based tests.

Method: An exploratory mixed methods design was used. Participants were 46 individuals with English as a first language (L1) and 23 participants with English as a second language (L2).

Results: The L1 and L2 participants differed only on the Language index of the SPANS, with L1 participants scoring significantly higher. Participants whose first language is not English reported generally that the test experience was good and the language used in the testing was clear. They reported some anxiety and worry in relation to memory and numbers.

Discussion: The results of this study support previous research which suggests that language may influence performance on language based neuropsychological tests as well as support for a need for culturally familiar test-items.

Conclusion: As there were minimal differences between the two groups support is provided for the SPANS's use with the tested population with caution in regard to interpretation of the language index.

Key words: Neuropsychological test, Language, Culture, SPANS

Exploration of the impact of completing the Short Parallel Test of Neuropsychological Status (SPANS) in a second language.

1. Introduction

The impact of cultural differences on neuropsychological test scores is not well understood. Expert opinion and limited empirical evidence provides a mixed picture of the effect. With an increase in multicultural societies in the UK it is important to better understand the impact on test scores of both linguistic factors, such as bilingualism, and cultural familiarity with the host culture (Lezak, 2012). Consideration as to whether it is appropriate to use a particular neuropsychological test with an individual whose culture-of-origin is different from that of the ‘culture’ of the test and its standardisation sample is a debated area with the extent of and application of such decisions varying widely (Cockcroft, Alloway, Copello & Milligan, 2015). Normative data is often developed in Western monolingual English-speaking countries, which is then used to assess performance with individuals from different cultures who speak English as a second language. Issues relating to this have been researched extensively in South Africa due to the number of languages spoken and cultural variations within the country. There are 11 different languages spoken with cultural variability, particularly in relation to socio-economic status, across the country. In areas of South Africa where language and culture are similar to Western Europe, IQ performance has been found to be comparable to normative samples (for example, Claasson, Krynauw, Paterson & Mathe, 2001). However, this is not the case across the whole country, and applying UK or US normative data without regard for linguistic or cultural implications could produce misleading results (Cockcroft, et al. 2015).

Cultural factors pertaining to the level and quality of education have been demonstrated to be important factors when completing neuropsychological tests in a second language and within a different culture to where the test was developed. Research exploring the impact of educational factors has been completed in South Africa where certain neuropsychological tests have culturally-validated item-administration procedures, for example on the WAIS-III (Claasson, et al. 2001). The results of two studies looking at the impact of both language and educational factors on the WAIS-III suggest that the Digit Symbol Coding task, when taken alone, does not appear to be impacted upon by language or education when comparing English first language (L1) and English second language (L2) participants in South Africa (Shuttleworth-Edwards, Donnelly, Reid & Radloff, 2004a). However, when considering the WAIS-III in its entirety both language and education impacted on South African L1 and L2 participants scores (Shuttleworth-Edwards, Kemp, Rust, Muirhead, Hartman & Radloff, 2004b). In particular, the quality of education was found to impact on both performance and verbal intelligence quotients (PIQ and VIQ respectively), suggesting an interaction between language ability and cultural factors, such as the acquisition of crystallised intelligence or test-taking skills.

Cultural familiarity with test-items has been shown to be an important factor when testing L2 individuals who originate from a different culture to where the test was developed. When test-items are adapted to be more culturally salient for the individual taking the test, fewer differences between L1 and L2 individuals have been observed. Adapting the Boston Naming Test (BNT) and the Cookie Theft Test (CTTest) to make test-items more culturally relevant for a South African population led to L2 participants scoring more in-line with normative data. With the original versions of the test, particularly the BNT, L2 performance was significantly below matched L1 controls. Additionally, participants reported high levels of unfamiliarity with the test-items (Mosdell, Balchin & Ameen, 2010). Other research looking at shopping-list verbal-memory assessments found that cultural-familiarity with test-items, along with delivering in the participant’s native language, were the most important factors for reducing differences between linguistic and cultural groups (Lim, Prang, Cysique, Pietrzak, Snyder & Maruff, 2009).

Direct translation of tests has been shown to be problematic if cultural considerations are not taken into account, and where a number of adaptations to the test are necessary, the validity of the test can be adversely affected (Renteria, Li & Pliskin, 2008; Tollman & Msenga, 1990). In some cases even when cultural adaptations for the translated test are made, performance between L1 and L2 matched groups have been found to still be significantly different (Demsky, Mittenberg, Quintar, Katell, & Golden, 1998). This suggests that careful consideration is required when using translated tests as well as a need for normative data for specific linguistic and cultural groups.

Research investigating the impact of completing a neuropsychological test in a second language has revealed that there are significant differences in performance between those with English as a second language (L2) and those speaking English as a first language (L1). This includes both language-based tests and those traditionally thought to be ‘culture-free’. For example, mean test score differences have been found using the Trail Making Test (TMT), a verbal learning memory test, the Wechsler Adult Intelligence Scale – revised (WAIS-R), Digit Span, Digit Symbol Coding, Stroop and Auditory Consonant Trigrams (ACT) tasks (Carstairs, Myors, Shores & Fogarty, 2006; Dugbartey, Townes & Mahurin, 2000; Harris, Cullum & Puente, 1995; Lu & Bigler, 2000; Razani, Burciaga, Madore & Wong, 2007). Assumptions are often made about L2 individuals’ ability to complete neuropsychological tests developed in Western countries and administered in English. When English-language ability is not taken into account prior to assessment, lower scores on tests which are a result of differences pertaining to language or culture could result in L2 individuals being classified as cognitively impaired when they are not. Assumptions may also be made about an individual's ability to speak and understand English proficiently. If their ability is overestimated L2 individuals could be at a disadvantage. This has been observed with more complex components of certain tasks, for example Stroop – B, TMT - B, ACT – 18s delay (Razani, et al. 2007).

Aims

The current study aimed to investigate the impact of language and culture on a recently developed neuropsychological test in the UK, the Short Parallel Assessments of Neuropsychological Status (SPANS: Burgess, 2014). Participants with English as a second language (L2) were compared to a sample of previously collected data from native English-speaking participants (L1). All participants' culture of origin was Western, though L2 participants all completed their early education and acquisition of English within a culture outside of the UK, and in a language other than English. There is little information in the literature pertaining to the experiences of L2 individuals completing neuropsychological assessments in a second language. Little is known about their perceived levels of proficiency in doing such assessments which could increase anxiety levels and as such negatively impact on their performance on the test (Lezak et al. 2004). This study therefore additionally aimed to qualitatively examine the experiences of L2 individuals completing a neuropsychological test in a second language.

Hypothesis

There will be differences between L2 and L1 participants on indices on the SPANS which are likely to be affected by language or issues related to cultural familiarity (for example naming, reading and writing a sentence).

Research questions

a) How do L2 participants find the experience of being tested on the SPANS in their second language?

b) Are there considerations that neuropsychologists need to take into account when testing L2 participants using the SPANS?

c) How might the experiences of being tested on the SPANS in a second language impact on differences between L2 and L1 subjects?

2. Method

2.1 Participants

Recruitment took place between January 2015 and March 2016. Initial recruitment began within a specific language subset; participants with a first language of South Asian origin (for example Bengali, Urdu, Hindi, Guajarati). This was due to this language subset constituting the highest non-European language subset in England and Wales (Census, 2011). Contact was made with various religious organisations, community groups, local centres and universities. However, there was a significant lack of response from the community within this language subset and no groups or individuals agreed to participate. A decision was then made to extend recruitment to anyone with a first language that is not English. Participants were recruited via word of mouth and advertisements in local libraries and universities.

Thirty-one L2 participants were recruited to the study. A total of eight participants were unable to complete the study. Reasons for drop-out related to the length of time required to complete the assessments (approximately 1-1.5 hours). A total number of 23 participants completed all quantitative aspects of the study and 15 of those completed the qualitative component. Eight participants were unable to complete the interview due to time restrictions. All participants were screened via questionnaire for the presence of any previous neurological disorder or trauma, as well as for the presence of other conditions which may have impacted on their scores on a neuropsychological assessment (Appendix 3).

English second-language data were compared to previously collected English first-language speakers (Burgess, 2014) data at a ratio of 2:1 (N=46) within the same age range. These data were collected as part of the original SPANS standardisation procedure.

2.2 Materials

SPANS: The SPANS is a short neuropsychological assessment battery that was originally developed for the assessment of neuropsychological status for inpatients following acquired brain injury (ABI). The SPANS was additionally developed so that it could be used in community outpatient clinics with those for whom an extensive battery of assessments would not be tolerated. The test was developed to be sensitive to impairment that may occur as a result of ABI without placing high demands on the patient in terms of extensive time or need for test-taking skills.

The SPANS has recently been published and has reliability and validity studies in preparation for publication. Pre-publication data has revealed that with L1 clinical and norming participants the SPANS evidenced satisfactory to excellent internal consistency (Table 1), as well as demonstrating that there is unique variance between the index scores. Adequate to excellent test-retest reliability was evidenced (Table 2) across the seven index scores (i.e. orientation, attention/concentration, language, memory/learning, visuo-motor performance, efficiency, and conceptual flexibility) on the SPANS. To assess inter-rater reliability, two subtests (Figure Copy and Similarities) were rated by five trained raters with a correlation coefficient of .95 agreement on the Figure Copy subtest and 100% agreement on the Similarities subtest.

Table 1

SPANS Internal Consistency (Cronbach Alphas)

	SPANS A	**SPANS B**	**Combined**
Orientation Index	.80	.77	.79
Attention/Concentration Index	.81	.84	.82
Language Index	.84	.88	.86
Memory and Learning Index	.89	.91	.90
Visuo-motor Performance Index	.85	.86	.85
Efficiency Index	.82	.86	.84
Conceptual Flexibility Index	.70	.77	.73

Table 2

Test-retest Reliability Coefficients for SPANS Indices (Cronbach Alphas)

	n = 26
Orientation Index	.85
Attention/Concentration Index	.92
Language Index	.97
Memory and Learning Index	.95
Visuo-motor Performance Index	.96
Efficiency Index	.93
Conceptual Flexibility Index	.86

In validity studies the SPANS significantly correlated with existing tests such as the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III; Wechsler, 1997a), the Wechsler Memory Scale, Third Edition (WMS-III; Wechsler, 1997b), the Rey Complex Figure (RCF; Meyers & Meyers, 1995) and the Trail-Making Test, Parts A and B (TMT; Reitan & Wolfson, 1985) demonstrating construct validity. The SPANS significantly correlated with tests known to have predictive ability such as the Performance IQ on the WAIS-III and the TMT-B.

In order to assess divergent validity, that is whether the SPANS tested for neuropsychological constructs that are impacted upon by ABI, the SPANS index scores were compared to the WAIS-III information subtest, which measures crystallised intelligence, and the Hospital Anxiety and Depression Scales (HADS: Zigmond & Snaith, 1983). As the SPANS is not intended to measure crystallised intelligence nor be a measure of mood or anxiety, limited correlation with these was expected. Overall the SPANS demonstrated divergent validity; however, there were small significant correlations between the WAIS Information subtest and the Attention Concentration Index (ACI) and Language Index (LAI) on the SPANS. The HADS anxiety scale correlated with the Visuo-motor Performance Index (VPI) and Memory Learning Index (MLI). These findings are discussed in relation to the skill sets that are shared across the measures, such as the academic skills required for both the Information subtest and listening and numeracy skills required for the SPANS. The relationship with the HADS anxiety scale is discussed in relation to anxiety related vigilance or conscientiousness for completing tasks such as Figures recall.

The SPANS was additionally found to have discriminant validity, differentiating between clinical (ABI less than one year and long-term neurological conditions) and healthy sample populations, indicating good sensitivity and specificity of the index scores. Those subtests that pertain to measure memory and performance related skills have higher discriminant power than those with language components, which is consistent with theoretical expectations (Burgess, 2014).

***Raven*'s Standard Progressive Matrices (SPM):** The SPM tests inductive reasoning (Alderton & Larson, 1990) and general intelligence (Raven, Raven & Court, 2000). It has also been linked to spatial and verbal reasoning ability (Lynn, Allik & Irwing 2004), working memory (Carpenter, Just & Shell, 1990) and speed of processing (Bates & Rock, 2004). The SPM has been demonstrated to have comparable psychometric properties across different cultural groups (McCallum, Bracken & Wasserman, 2000; Rushton, Skuy, & Bons, 2004) and was therefore used as an additional screening measure within this study. Any participant with a SPM score within the impaired range (below the 2nd percentile) would not be included in the study as their overall level of functioning may be a confounding factor for scores on the SPANS.

An exploration of the reliability and validity of the SPM has found good internal consistency reliability (r= .88). The SPM has been reported as having construct validity in terms of assessing analytic intelligence (Carpenter, Just & Shell, 1990) and educative ability (Raven, Court & Raven, 1998). The SPM has been shown to have adequate convergent validity, correlating with scales of the WAIS-III (Matrix reasoning r =.81; Performance IQ r=.79; Full scale IQ r=.64 and verbal IQ r=.49) (Wechsler, 1997a).

Language History Questionnaire (LHQ): The LHQ (Li, Zhang, Tsai & Puls, 2013) [Appendix 8] was used to assess participants' second language proficiency. This allowed for the evaluation of the number of years spent speaking English as a second language, as well as gaining an understanding of the participant's proficiency in reading, writing, speaking and understanding English as a second language. Adequate split-half reliability (r=.85), construct validity (age of L2 acquisition, number of years of learning, amount of L2 use, self-assessed reading, speaking and comprehension) and predictive validity (prediction of L2 proficiency) have been reported (Sepanski, 2005; Sepanski & Li, 2006).

2.3 Design and procedures

This study adopted an exploratory mixed methods design. For the quantitative component of the study an independent (between subjects) design was used; comparing group means (index scores & individual subtests) of English second-language speakers (L2) and English first-language speakers (L1). For the qualitative component of the study content analysis (Elo & Kyngäs, 2008) of a semi structured interview (Appendix 2) was conducted to explore the experiences of the participants being assessed in their second language.

Data Analysis: An initial sample size of 38 was anticipated in order to detect a large effect size of .01 with 80% power (Cohen, 1992). Statistical analyses were performed using SPSS v23. Mann-Whitney U-tests were performed to determine whether there were significant differences between L1 and L2 participants on each index score. The two-tailed Mann-Whitney test was chosen to correct for violations relating to the normality assumption of the data. Tests were two-tailed and the alpha level for statistical significance was set at .01 to reduce the risk of Type 1 error.

Inductive qualitative content analysis was used to categorise the data collected from the semi-structured interviews following the recommendations outlined by Elo and Kyngäs (2008). Open coding was used to create sub- and generic-categories which were then abstracted into main categories. Credibility of the qualitative data was checked using the Kappa measure of agreement. The Kappa value over two ratings of the coded data was Kappa = 1.00 ($p = .000$) indicating “very good” agreement (Appendix 4).

Ethical considerations: Ethical approval was received from the Canterbury Christ Church University Research Ethics Committee (Appendix 5). Participants were given the opportunity to meet in areas local to them to minimise the impact of the time taken to complete the assessments. All participants were asked to complete a consent form (Appendix 6) and were provided with an information sheet (Appendix 7) outlining the objectives of the study. All participants were asked to be able to speak conversationally in English in order to complete the neuropsychological assessments, therefore the information sheet and consent form were also provided in English. A participant ID was given to all to protect anonymity. Participant data will be held for ten years following completion of the study. The testing situation was not anticipated to cause any significant distress or harm to those participating. Details in the information sheet advised that individual test results would not be made available as they relate to research rather than representing a clinical measure of functioning. Participants were given researchers' contact information and advised that they could be approached if there were any concerns about the tests or results. British Psychological Society (BPS) code of ethics and conduct was adhered to at all times throughout the project (BPS, 2011).

3. Results

3.1 Quantitative results: Participant characteristics

A total of 69 participants were included in the quantitative analysis (L1 = 46, L2 = 23). Independent sample t-tests revealed that there were no significant differences between L1 (M = 25.22, SE = .84) and L2 (M = 28.30, SE = .92) participants in terms of age, $t(67) = -2.27$, $p = .026$. L1 participants were 43% male and 57% female and the L2 participants 52% male and 48% female. With regard to level of education, 26% of L1 participants had higher level education, 41% college and 33% secondary. For the L2 participants 61% had higher level education, 26% college and 13% secondary.

The L2 participants originated from 12 different nationalities all within Western European culture (Table 3). They had an average of 19 years speaking English (M = 19.43, SD = 4.75) and spoke an average of 3 languages (M = 3.13, SD = .76). With regard to perceived proficiency in listening, speaking, reading and writing in English all reported "very good" perceived ability, indicated by a score of 6 on the LHQ (Table 4). All L2 participants scored above the 2nd percentile on the SPM and were classified as falling within the average-high average ranges of general intelligence and were therefore all included in the subsequent analyses.

Table 3

Frequencies of Nationalities in L2 Sample

	Frequency in L2 Sample
Polish	2
Czech	2
Dutch	6
Spanish	2
Portuguese	1
Italian	2
Turkish	1
Greek	1
Albanian	1
Swedish	1
French	2
German	2

Table 4

L2 Perceived English Ability (LHQ)

	M (SD)
Perceived ability in listening in English	6.04 (.77)
Perceived ability in speaking in English	5.91 (.73)
Perceived ability in reading in English	6.09 (.73)
Perceived ability in writing in English	5.96 (.77)

Note: Range of scores 1 (very poor) to 7 (native-like) A score of 6 represents *"very good"* perceived ability

3.2 Quantitative Analysis

The means and standard deviations for both groups on each of the SPANS index scores are presented in Table 5.

Table 5

L1 and L2 SPANS Index T-Scores

	L1 **M (SD)**	**L2** **M (SD)**
Orientation Index	49.52 (8.60)	40.09 (13.79)
Attention/Concentration Index	50.15 (11.28)	47.30 (11.68)
Language Index	50.61 (9.70)	39.04 (11.98)
Memory and Learning Index	46.13 (12.47)	45.83 (9.90)
Visuo-motor Performance Index	49.43 (11.21)	53.17 (7.02)
Efficiency Index	49.96 (11.56)	50.13 (8.42)
Conceptual Flexibility Index	49.91 (8.58)	48.09 (7.95)

Note: Maximum T-score for each index (age range 18-50) are as follows: ORI = 53, ACI = 64, LAI = 63, MLI = 68, VPI = 68, EFI = 61, CFI = 56

The L1 group significantly outperformed the L2 group on the Orientation Index (ORI): L1 participants (Mdn = 53) and L2 participants (Mdn = 53), $U = 706.50$, $Z = 2.88$, $p < .01$, $r = .35$ and Language Index (LAI): L1 participants (Mdn = 54) and L2 participants (Mdn = 41), $U = 809.50$, $Z = 3.61$, $p < .01$, $r = .43$ only.

There were no differences between the groups for the Attention Concentration Index (ACI): L1 participants (Mdn = 53) and L2 participants (Mdn = 49), $U = 603.50$, $Z = .95$, $p = .340$, $r = .11$; Memory Learning Index (MLI): L1 participants (Mdn = 50) and L2 participants (Mdn = 45), $U = 590.00$, $Z = .79$, $p = .433$, $r = .10$; Visuo-motor Performance Index (VPI): L1 participants (Mdn = 49) and L2 participants (Mdn = 55), $U = 444.00$, $Z = -1.10$, $p = .270$, $r = 0.13$; Efficiency Index (EFI): L1 participants (Mdn = 51) and L2 participants (Mdn = 51), U =

572.00, Z = .55, p = .579, r = .07 or the Conceptual Flexibility Index (CFI): L1 participants (Mdn = 54) and L2 participants (Mdn = 54), U = 622.00, Z = 1.29, p = .197, r = .16.

Due to the small sample size and increased risk of Type 1 error due to multiple comparisons, a Bonferroni correction was applied. With this correction the new familywise error rate to detect statistical significance was .001. Using this level of significance the LAI index produced the only significant difference between the two groups, with the L1 group (L1: M = 50.61, SD = 9.70) achieving significantly higher index scores than the L2 group (M = 39.04, SD = 11.96). The bootstrapping process was computed using SPSS as a robust resampling methodology to eliminate the need for normally distributed data. Using this resampling method did not alter the significant results from the Mann-Whitney U Tests and is thus not reported on further.

In order to identify which particular LAI subtests produced significant differences between L1 and L2 participants, Mann-Whitney U tests were performed on all LAI subtests. The L1 group scored significantly higher on the repetition subtest (Mdn = 6) than the L2 group (Mdn = 5), U = 806.00, Z = 3.93, p < .001, r = .47. The L1 group additionally scored significantly higher on the naming subtest (Mdn = 12) than the L2 group (Mdn = 10), U = 874.00, Z = 6.11, p < .001, r = .74. Surprisingly, the L2 group scored significantly higher on the writing sentences subtest (Mdn = 5) than the L1 group (Mdn = 4), U = 298.00, Z = -3.44, p < .001, r = .41.

There were no significant differences between the L1 and L2 participants for the Yes/No Question subtest (L1 Mdn = 6, L2 Mdn = 6), U = 529.00, Z = .000, p = 1.00, r = 0; the Following Directions subtest (L1 Mdn = 6, L2 Mdn = 6), U = 460.00, Z = -1.51, p = .132, r = .18; the Reading subtest, (L1 Mdn = 6, L2 Mdn = 6), U = 517.50, Z = -.71, p = .480, r = .09; or the Similarities subtest (L1 Mdn = 12, L2 Mdn = 12), U = 630.00, Z = 1.61, p < .107, r = .19.

3.3 Qualitative Results

Fifteen participants were interviewed about their experiences of being tested on the SPANS in their second language using a semi-structured interview process. These data were analysed using qualitative content analysis. The coding process resulted in a total of sixteen sub-categories being generated which were abstracted further into three generic categories. A description of each of the generic categories with quotes from transcripts is given below. A diagrammatic example of the abstraction process, example quotes and a sample transcript are included in the appendices (9 10 and 11). Table 6 gives the percentage and number of participants providing responses to the questions in the semi-structured interview.

Table 6

Qualitative responses from L2 participants regarding the SPANS

	Yes **% (n)**	**No** **% (n)**
Was the language accessible?	87% (13)	13% (2)
Did the participant feel disadvantaged?	60% (9)	40% (6)
Could anything make the testing experience easier?	20% (3)	80% (12)
Would the participant like anything to be different?	13% (2)	87% (13)
Would the participant change anything about the instructions?	13% (2)	87% (13)
Did the participant feel anxiety doing the SPANS?	60% (9)	40% (6)
Did the participant feel worry doing the SPANS?	53% (8)	47% (7)
Was there anything that surprised the participant when doing the SPANS?	47% (7)	53% (8)

3.4 Positive and negative experiences of the test

Language used to describe the experience of the SPANS consisted of generally positive descriptions. There was a sense that the experience was somewhat enjoyable, with 60% of participants using language that portrayed enjoyment and that the overall experience of being tested was positive.

"*L003: the SPANS yeah, umm, it was…fun, if that makes sense, I thought* you know, the, the particularly the symbol shape things, I just thought that was quite interactive, so it is like learning a skill almost and yeah I found it quite interesting that I had some different *thoughts…like the banana and yellow…"*

As the majority of the sample were educated to degree level this may have contributed to their interest in the testing experience and their ability to find being tested enjoyable. Additionally, as each of the SPANS subtests is quite short and the tasks are varied, this may have also contributed to the enjoyment of the task. Those who had experiences of being tested with psychometric measures in the past described the test as "refreshing" and "not like a test". This adds to the desirability of the test for use with clinical populations, especially those who have an ABI whose ability to manage demands of an extensive neuropsychological battery may be compromised.

Some of the sample (47%) felt that the test at times was "challenging" and "complicated" and for a minority (20%), that at times the questions were so easy that it was potentially "insulting". Some found it difficult in relation to anxieties relating to remembering and the use of numbers, though overall most reported satisfaction with their experience of the test.

Over half of the interviewed L1 participants reported that the test situation led them to feel worried (53%) or anxious (60%) about their ability to complete the test. This appeared to relate to concerns about getting the question right or their ability in specific areas such as drawing (Figure Copy).

"L028: err, well just a little bit about drawing because I'm not very good at drawing…"

"L027: …well, I dunno, cos I just, I wanted to give the right answer but, it was hard to, yeah"

"R: Were there any specific parts of the test that caused you to feel anxious?

L018: umm…being wrong but that's common isn't it?"

3.5 Feelings associated with memory and language

87% of interviewed participants reported that the language used in the testing was accessible. For the two participants who did not find it accessible this related to not knowing one of the words used in the test and confusion over the structure of a question relating to money (Monetary Calculations). The instructions were generally described as "clear" and 87% of interviewed participants would not change anything about them. For those who would change something, there was a sense that certain parts could be more straightforward, which again linked to some confusion over a Monetary Calculation question and whether the answer was £100, or £35 ("And if you were given £65 change after spending £100, how much did you spend").

Worry about remembering was a key theme with the participants interviewed, with 67% of participants indicating this. There was a sense from one participant that the level of anxiety about their memory made them worry that they had a memory problem such as dementia. Others worried that they were unable to remember small amounts of information and that this may mean they had a bad memory.

"L011: *umm…I was quite nervous to start with…then I couldn't work out if I got dementia (laugh), because I couldn't remember stuff"*

"L022: how easy it is to forget even when someone has read to you for numbers, you can just forget it"

Some worried about their perceived ability to speak English adequately, for example having the correct vocabulary to complete the Naming subtest and it was felt by one that the familiarity of the pictures on the Naming subtest may have contributed to this.

"L011:…I'm thinking about the four, or five, or however many pictures, and obviously I didn't happen to know two of them…and they're fairly common words but not as common as like a cat or car or something…so it might be helpful to have something that is more simpler"

"…I mean apart from my lack of vocabulary for, what was that tool again…"

3.6 Feelings associated with numbers

Issues relating to numbers were discussed by 73% of the participants both in relation to worry and anxiety. This appeared to be in relation to participants' perceived ability to do numerical calculations, as well as relating to the ability to hold in mind numbers that were heard in their second language but processed in their native language. This was reported to increase the demands on the Digits-Forward and Backwards-subtests in particular. For those who reported that numbers were still processed in their native language, number tasks such as Digits Forwards and Backwards as well as Monetary Calculations and Counting Backwards were reported to be anxiety provoking. There was a sense that performance on tasks relating to numbers stayed with the participants with one making reference to poor performance on number tasks a total of three times throughout the interview.

"R: Were there any specific parts of the test that caused you to feel anxious?

L014: counting backwards in 3's"

"L014: …how bad I was at counting backwards"

"L002: …I was just thinking with the number sequence how I would have done if it had been in Czech".

*"R: Were there any specific parts o*f the test that caused you to feel anxious?

L022: umm…the math with the adding and the umm…thing, minus and plus"

3.7 Potential impact of experiences of being tested in a second language on the differences between L1 and L2 participants

The quantitative results of this study suggest that L2 participants are more likely to score lower than L1 participants on the Language Index of the SPANS. This links to the qualitative reports from L2 participants that worries about their vocabulary were present during the testing situation and that this impacted upon their perceived ability to do the test. This is noteworthy given that the L2 participants perceived ability to speak, read, listen and write in English was rated as "very good". Although participants reported that they worried about their number ability, this did not appear to impact upon their performance on tasks involving numbers as there were no significant differences between the groups on indices involving numerical tasks.

4. Discussion

4.1 Impact of language and culture on the SPANS

The current study aimed to explore whether there were differences between the mean index scores on the SPANS with people whose first language was not English compared with those whose first language was English. The main finding was that the Language Index of the SPANS was the only area with a significantly lower index score for L2 participants in this sample. Additionally, the experiences of the L2 participants being tested in their second language on the SPANS was explored; the findings suggesting that even those whose perceived ability to speak, read, write and listen in English was ”very good”, worried about their ability when in a situation which was perceived to be testing their language skills.

This exploratory study demonstrates similarities with previous research which has examined the differences between L1 and L2 participants, finding that language-based tasks show differences between the two groups (e.g. Carstairs et al. 2006; Harris et al. 1995) but not on more performance-based tasks, such as on tests of information processing and attention (e.g. Razani et al. 2007) and the TMT-B (Lu & Bigler, 2000). This adds support for the SPANS in terms of its cultural fairness; the majority of the tasks were understood and interpreted correctly by L2 participants and overall the interviews conducted did not reveal that participants were unfamiliar with the test-items. This was only the case with one of the objects within the Naming subtest, with 65% of L2 participants not knowing the name of at least one of the objects.

The least correctly-named item on the Naming subtest was the "spanner", with a number of participants remarking that they had never heard of the English word for this object. This suggests that this is an object that does not have the same level of salience across cultures, even those which share similar characteristics such as the Western European countries in the current study. Cultural familiarity with test-items has been reported as an important factor for minimising the impact of linguistic and cultural bias when using tests developed in one language, with L2 participants (Lezak, 2012; Mosdell, et al. 2010; Lim, et al. 2009). When using the SPANS with those whose first language is not English it may be necessary to use a different object in place of the spanner, in order to minimise the risk of a lower score being a result of a lack of familiarity with an object, rather than a problem with language per se. Consideration should also be given to ensuring that not knowing the English name of an item such as "spanner" does not impact on the memory recall component of the test. Previous research has demonstrated that differences on memory test scores between L1 and L2 groups is more likely to be a result of not being tested in a participant's dominant language rather than problems with memory (Harris et al, 1995).

The other language subtest which was lower for the L2 group was the Repetition subtest. This task requires the subject to repeat a number of phrases with differing levels of complexity. Due to the high demands on language in terms of pronunciation and sentence structure it is unsurprising that L2 participant's performance on this subtest was lower than that of the L1 participants. Participants were observed to consistently alter the sentence structure of the phrase "the size of the house was one-tenth the size of the mansion" adding in an additional "of" after "one-tenth". This may relate to their native sentence structure which is likely to be different to that of the English-language. Similar findings have been observed even within the same language where certain letters are pronounced differently dependent on the region the language is spoken in (Renteria et al. 2008).

Surprisingly, the L2 participants performed the Writing Sentences subtest better than L1 participants. This may relate to the L2 participants learning English more recently than L1 participants and therefore they may be more consciously aware of sentence structure and grammar. The difference between the groups, though significant was only a difference of 1-point and could also be a consequence of being marked by two different people. However, the SPANS has been demonstrated to have excellent inter-rater reliability (Burgess, 2014) and as such differences in marking would be not be expected.

Previous research has demonstrated that level and quality of education may be a factor contributing to differences in L1 and L2 performance on tests such as the WAIS-III (for example Shuttleworth, et al. 2004b). In the current study, although quality of education was not specifically measured, the level of educational achievement was obtained for each participant in both the L1 and L2 groups. The L2 participants were more likely to have had a higher-level qualification than the L1 groups demonstrated by a higher percentage of L2 participants completing their education to degree level. This could contribute to the lack of difference between the two groups, for example higher educational levels contributing to the L2 "normal" scores. Having more experience with taking tests at increasingly demanding levels may mean that they are at an advantage with regard to test-taking skills. Achieving a higher level of education may additionally have contributed to their ability to learn and be proficient in a second language, meaning that their level of English is higher than those who have a lower level of education.

Despite this, the fact that the L2 group were generally highly educated makes the finding that their LAI score was significantly lower than the L1 participants more clinically significant. Those with a high level of educational achievement, who have spoken English for over 10 years and whose perceived ability with the English language was rated as being "very good", would not be expected to score significantly below the L1 participants on relatively simple tasks. This suggests that language and familiarity with test-items on the language subtest, such as the "spanner", are likely contributors to this lower score. Additionally, the SPANS is not a measure of general intelligence and as such should not be impacted upon significantly by level of education (Burgess, 2014).

The entire L2 sample completed their education in their native country; that is outside of an English speaking country. Previous research in the United States of America (USA) has suggested that completing education outside of the USA can negatively impact scores on tests developed and standardised in the USA (Razani, et al. 2007). This is discussed in relation to the acquisition of test taking skills or "testwiseness", which is the ability to know the expectations required for taking a test; and 'tests' themselves are culturally-specific. Additionally, research looking at the quality of education, particularly in South Africa, has demonstrated this to be an influential factor contributing to differences between L1 and L2 groups (e.g. Shuttleworth, et al. 2004). In the current study this is unlikely to have had an influential role due to all of the participants' education being in Western European countries and generally to a high-level. Though there may be subtle differences in the educational systems, it is unlikely that the quality of education would have differed significantly.

Although participants reported that they worried about their abilities to hold and manipulate numerical information in mind due to complexities with translation, there were no differences between the L1 and L2 groups on indices with numerical tasks included. This suggests that the anxiety with regard to numbers reported by L2 participants in this study sample may not be related specifically to their ability to do the task. It may relate to the increased demands of the task, as well as anxiety related to confidence in numerical tasks, which may have stemmed from their educational experiences with Mathematics. High levels of test anxiety have been shown to negatively impact upon test results (for example Ganzer, 1968; Sarason, 1982), particularly in relation to cognitive interference (worry). Additionally when the achievement orienting conditions for the test are low that is the demand placed upon the test-taker to do well, anxiety about the test is lowered (Sarason, 1982). In the current study, the demands placed upon the participants with regard to performance expectations were low; there was no benefit or cost to them if they did poorly on the test, therefore anxiety levels would be anticipated to be low. Although anxiety in some form was reported by over half of the participants, the qualitative analysis did not reveal that this was at a level where participants felt it negatively impacted upon their cognitive ability to complete the test.

4.2 Limitations of the study

The sample for the L2 group was small and this contributes to low power for the study. This could have an impact on the significance of the results and the extent of the effect sizes seen, for example, the chance of finding a significant result when there is not one (a Type II error). Post-hoc corrections were computed to minimise the impact of this, however future research in this area with higher sample sizes would be beneficial to reduce this risk.

The sample used in this study was largely heterogeneous. This means that although the findings suggest that L2 participants originating from Western Europe are likely to score lower on the LAI as compared to L1 participants in the same age range, inferences about specific language groups cannot be made.

The current sample for L2 participants represents a small age range. This reduces the generalisability of the findings to the wider population and inferences from the findings are limited to the age range in the study sample. Additionally, as the majority of the L2 group were highly educated, the impact of this on the results cannot be ruled out and again impacts on the ability to generalise the findings to the population as a whole.

Difficulties with recruiting participants to the study reduced the variability in the study sample in relation to areas such as educational ability, time spent in the U.K and time spent speaking English. Future research may be able to target specific linguistic groups which for the purposes of the current study were difficult to reach in part due to time limitations.

A further limitation relates to the semi-structured interview. The people interviewed were asked to comment on their experience of a specific test knowing that the results contributed to a research study, and interviews were conducted by the same person who administered the test. This may have led to the participants speaking more favourably about the test and administration so as not to cause offence. However, participants did speak openly at certain points in the interview; therefore it is unlikely that the impact of demand characteristics was high enough to have produced a large effect.

The current study used a measure of language proficiency (LHQ) which was based on the participants' perception of their ability in speaking, reading, writing and listening in English. Overall, the sample group was rated as "very good". This again impacts on the ability to generalise the findings of the research to those who may not perceive their English-language ability to be as "good" as the study sample**.** Additionally this was self-reported ability and therefore may have been over reported as a result of social desirability bias.

4.3 Clinical and theoretical implications

Comprehensive neuropsychological assessment plays a key role in the understanding of the consequences of neurological disorders. Having a robust, reliable and valid measure for assessing the cognitive deficits associated with such disorders is vital for ensuring correct and timely diagnosis, as well as informing neurorehabilitation. With the increase in cross-cultural migration it is imperative that neuropsychological tests used routinely within the UK are valid for the population they are testing.

The results of the current study suggest that there is a need for clinicians using the SPANS with people whose first language is not English (L2), to be cautious when interpreting the Language Index (LAI). The current study suggests that the Naming and Repetition subtests from the LAI are the most likely to be scored lower by L2 participants, and particularly with the current sample, participants were unlikely to know the name of a "spanner". With regard to the

Repetition subtest, differences with sentence structure may be observed with words repeated in different orders or additional words added to the phrases. As these subtests were found to contribute to a significant difference in the overall index score, care needs to be taken when scoring such tasks with L2 individuals to ensure that they are not scored unfavourably due to language proficiency or cultural familiarity, as opposed to the construct the index is designed to measure (language ability). Checking whether the individual knows what a spanner is used for and if they know the name in their native language may be useful for identifying whether the difficulty is with naming or English-language proficiency.

4.4 Future research

Future research with the SPANS could explore the use of the test with people who have been speaking English for less than ten years to examine whether there are differences between L1 and L2 participants when the number of years speaking English is less. It may be that this research would need to be with people from a non-Western background as many Western European cultures learn English during their school years.

5. Conclusion

This exploratory study has demonstrated that the SPANS is generally a robust neuropsychological test that can be used with young adults of Western European origin who have had significant experience (average of 19 years) speaking English. Caution needs to be applied when interpreting performance on the Language Index as the results of the current study indicate lower than expected scores with this population, probably due to differences in language structure and familiarity with one or two test-items. Within this study sample the impact of completing a test in a second language was reported to be minimal, although worries about perceived ability with numbers and vocabulary in particular was present for a number of participants. This may increase in clinical situations and could impact on their ability to complete the test if the expectations of the test-taking situation had a direct impact on the outcome of their treatment.

References

Alderton, D.L. & Larson, G.E. (1990). Dimensionality of Raven's Advanced Progressive Matrices items. Educational and Psychological Measurement, 50, 887-900.

Bates, T.C. & Rock, A. (2004). Personality and information processing speed: Independent influences on intelligent performance. Intelligence, 32, 33-46.

British Psychological Society (2011). Code of human research ethics. Leicester: BPS.

Burgess, GH. (2014). Short Parallel Tests of Neuropsychological Status: Test manual. Oxford, UK: Hogrefe.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. Psychological Review, 97(3), 404-431. doi: 10.1037/0033-295X.97.3.404

Carstairs, J.R., Myors, B., Shores, E.A., & Fogarty, G. (2006). Influence of language background on tests of cognitive abilities: Australian data, Australian Psychologist, 41(1): 48-54.

Census (2011). Local characteristics: main language by sex by age. Retrieved from http://www.nomisweb.co.uk/census/2011/LC2104EW/view/2092957699?rows=main_language&cols=c_age

Claassen, N. C. W., Krynauw, A. H., Paterson, H., and Mathe, M. (2001). A Standardisation of the WAIS-III for English-Speaking South Africans. Pretoria: Human Sciences Research Council.

Cockcroft, K., Alloway, T., Copello, E., & Milligan, R. (2015). A cross-cultural comparison between South African and British students on the Wechsler Adult Intelligence Scales Third Edition (WAIS-III). Frontiers in Psychology, 6:297. doi: 10.3389/fpsyg.2015.00297

Cohen, J. (1992). A Power Primer, Psychology Bulletin, 112 (1) 155-159

Demsky, Y.I., Mittenberg, W., Quintar, B., Katell, A.D., & Golden C.J. (1998). Bias in the use of standard American norms with Spanish translations of the Wechsler Memory Scale-Revised. Test. 5(2), 115-121.

Dugbartey, A.T., Townes, B.D., & Mahurin, R.K. (2000). Equivalence of the Colour Trails Test and Trail Making Test in non-native English-speakers. Archives of Clinical Neuropsychology, 15 (5), 425-431.

Elo, S. & Kyngäs, H. (2008). The qualitative content analysis process. Journal of Advanced Nursing, 62 (1), 107-115 doi: 10.1111/j.1365-2648.2007.04569.x

Ganzer, V. J. (1968). Effects of audience presence and test anxiety on learning and retention in a serial learning situation. Journal of Personality and Social Psychology, 8 (2p1), 194.

Harris, J.G., Cullum, M., & Puente, A.E. (1995). Effects of bilingualism on verbal learning and memory in Hispanic adults, Journal of International Neuropsychological Society, 1, 10-16.

Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). Neuropsychological test (4th ed.). New York, NY: Oxford University Press.

Lezak, M. D. (2012). Neuropsychological test. Oxford: Oxford University Press.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language History Questionnaire (LHQ 2.0): A new dynamic web-based research tool. Bilingualism: Language and Cognition, 17(3), 673-680. doi:10.1017/S1366728913000606

Lim, Y. Y., Prang, K. H., Cysique, L., Pietrzak, R. H., Snyder, P. J., & Maruff, P. (2009). A method for cross-cultural adaptation of a verbal memory test. Behavior Research Methods, 41(4), 1190-1200.

Lu, L. & Bigler, E.D. (2000). Performance on original and a Chinese version of Trail Making Test Part B: A normative bilingual sample, Applied Neuropsychology, 7:4, 243-246. DOI: 10.1207/S15324826AN0704_6

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

McCallum, S., Bracken, B., & Wasserman, J. (2000). Essentials of non-verbal test. New York: John Wiley & Sons.

Meyers, J.E. & Meyers, K.R. (1995). Rey Complex Figure Test and Recognition Trial. Lutz, FL: Psychological Test Resources.

Mosdell, J. Balchin, R. & Ameen, O. (2010). Adaptation of aphasia tests for neurocognitive screening in South Africa. South African Journal of Psychology, 40 (3), 250-261.

Raven, J. C., Court, J. H., & Raven, J. (1998). Raven Manual: Section 1, general overview, 1998 edition. Oxford, UK. Oxford Psychologists Press ltd.

Raven, J. Raven, J.C., & Court, J.H. (2000). Raven manual: Section 3. Standard Progressive Matrices. Oxford: Oxford Psychologists Press ltd.

Razani, J., Burciaga, J., Madore, M., & Wong, J. (2007). Effects of acculturation on tests of attention and information processing in an ethnically diverse group. Archives of Clinical Neuropsychology, 22, 333-341.

Reitan, R.M. & Wolfson, D. (1985). The Halstead-Reitan Neuropsychological Test Battery. Tucson, AZ: Neuropsychology Press.

Renteria, L., Li, S. T., & Pliskin, N. H. (2008). Reliability and validity of the Spanish language Wechsler adult intelligence scale in a sample of American, urban, Spanish-speaking

Hispanics. The Clinical Neuropsychologist, 22(3), 455-470. DOI: 10.1080/13854040701336428

Rushton, J. P., Skuy, M., & Bons, T. A. (2004). Construct Validity of Raven's Advanced Progressive Matrices for African and Non-African Engineering Students in South Africa. International Journal of Selection and Test, 12 (3), 220-229.

Sarason, I.G. (1982). Stress, anxiety, and cognitive interference: Reaction to tests. Journal of personality and social psychology, 46(4), 929.

Sepanski, S. E. (2005). Sentence recognition in native and foreign languages: Comprehension of form and meaning. Unpublished MA thesis, University of Richmond.

Sepanski, S. E., & Li, P. (2006). Validation of a generic language history questionnaire for L2 research. Unpublished manuscript, University of Richmond.

Shuttleworth-Edwards, A.B., Donnelly, M.J., Reid, I., & Radloff, S.E. (2004a). A Cross-cultural study with culture fair normative indications on WAIS-III Digit Symbol—Incidental Learning. Journal of Clinical and Experimental Neuropsychology, 26(7), 921-932. doi 1380-3395/04/2607-921

Shuttleworth-Edwards, A.B., Kemp, R.D., Rust, A.L., Muirhead, J.G.L., Hartman, N.P., & Radloff, S.E. (2004b) Cross cultural effects of IQ test performance: A review and preliminary normative indications on WAIS-III test performance. Journal of Clinical and Experimental Neuropsychology. 26: 7 DOI: 10.1080/13803390490510824

Tollman, S.G., & Msengana, N.B. (1990). Neuropsychological test: Problems in evaluating the higher mental functioning of Zulu-speaking people using traditional western techniques. South African Journal of Psychology, 20 (1).

Wechsler, D. (1997a). Wechsler Adult Intelligence Scale-Third Edition. London: The

Psychological Corporation Limited.

Wechsler, D. (1997b). Wechsler Memory Scale-Third Edition. London: The Psychological Corporation Limited.

Zigmond, A.S. & Snaith, R.P. (1983). The Hospital Anxiety and Depression Scale. Acta Psychiatrica Scandinavica, 67, 361–370.

Section C

Appendices

An exploration of the impact of language and culture on a neuropsychological test

September 2014

SALOMONS CANTERBURY CHRIST CHURCH UNIVERSITY

Appendix 1: Table of included studies (Section A)

Reference	**Language or Cultures**	**Neuropsychological Tests used**	**Sample Characteristics**	**Main Results (statistics reported when mentioned in results of the study)**
Cockcroft, K., Alloway, T., Copello, E., & Milligan, R. (2015). A cross-cultural comparison between South African and British students on the Wechsler Adult Intelligence Scales Third Edition (WAIS-III). Frontiers in Psychology, 6:297. doi: 10.3389/fpsyg.2015.00297	South African English	WAIS-III	**N=456** 349= undergraduate UK students 107= black African undergraduate students Age range: 18-58 years (UK) 18-25 years (African) Education: university level	MANOVA's with post-hoc bonferroni corrections (p<0.001) indicated that British students scored significantly higher than the SA students all verbal subtests and all non-verbal subtests except for Matrix Reasoning. MANOVA with post-hoc bonferroni corrections (p<0.001) indicated that the SA students scored higher than the British students on all PSI subtests.
Mosdell, J. Balchin, R. & Ameen, O. (2010). Adaptation of aphasia tests for neurocognitive screening in South Africa. South African Journal of Psychology, 40 (3), 250-261.	South African	Boston Naming Test (BNT) & culturally adapted version Groote Shuur Naming Test Cookie Theft Test (CTT) & culturally adapted version Washing Line Picture Test	**N=30** 10 English 10 Afrikaans 10 isiXhosa speakers Age range= 19-45 years Education= 7-13 years	All groups performed significantly better on the GSNT than the BNT: the **Afrikaans group** (t (9) = 8.81, p < .001; M(SD) 27.4 (1.65) versus 18.4 (3.84)), the **isiXhosa group** (t (9) = 9.58,p < ,001; M(SD) 26.8 (1.99) versus 13.1 (5.47)), and the **English group** (t (9) = 4.46, p < .01; M(SD) 27.9 (1.91) versus 20.4 (6.98)).

				The Washing line test was more familiar than the Cookie Theft Test (chi square $X^2(2, N = 30) = 9.6, p<.001$)
Lim, Y. Y., Prang, K. H., Cysique, L., Pietrzak, R. H., Snyder, P. J., & Maruff, P. (2009). A method for cross-cultural adaptation of a verbal memory test. Behavior Research Methods, 41(4), 1190-1200.	American (English) Australian (English) French Malaysian (Mandarin) (Malay)	Shopping list test (translated into appropriate languages)	**N=97** 25 American undergraduates Age range = 20-27 years 25 Australian undergraduates Age range = 19-27 years 25 French undergraduates Age range = 21- 26 years 11 Mandarin (no educational information) Age range = 18-35 years 11 Malay speakers (no educational information)	Exp 1: the use of culturally familiar words was the most important factor for increased memory performance Exp 2: highly culturally familiar words produced higher recall in both languages. Directly translated tests resulted in poorer performance. Exp 3: participants from the same culture but with different languages had equivalent performance on the test that used familiar words in the participants language Exp 4: bilingual performance - there were no significant differences in scores regardless of list (Mandarin or Malay) used

			Age range = 18-35 years	
Renteria, L., Li, S. T., & Pliskin, N. H. (2008). Reliability and validity of the Spanish language Wechsler adult intelligence scale in a sample of American, urban, Spanish-speaking Hispanics. The Clinical Neuropsychologist, 22(3), 455-470. DOI: 10.1080/13854040701336428	Spanish (USA) English	Measures of Intelligence (Spanish WAIS-III; TEA Ediciones, 2001 and Standard Progressive Matrices [SPM]; Raven, Court & Raven, 1992) Visuospatial abilities (Rey Complex Figures Test; Meyers and Meyers, 1995); Language (Controlled Oral Word Association Test [COWA]; Benton & Hamsher, 1976) Attention (Colour Trails 1 & 2 [CTT1 & CTT2]; D'Elia, Satz, Uchiyama, & White, 1994; Strauss, et al., 2006).	**N=100** 100 Spanish speaking ppts (50 male, 50 female) living in USA. Age range= 25-54 years Education= 1-18 years Performance compared to Spanish and American WAIS-III norms.	Adequate reliability on all subtests except letter-number sequencing (internal consistency X=.62) Evidence of criterion-related validity indicated by strong correlations between SPM and the FSIQ, PIQ & POI. However paired t-tests revealed significantly lower FSIQ and PIQ scores compared to the SPM ($p<.001$). Evidence of convergent-language-related validity indicated by strong correlations between COWA and VCI as well as VIQ (though lower than anticipated). Paired t-tests revealed significantly lower VCI & VIQ scores compared to the COWA ($p<.001$). Evidence of convergent-visuospatial-related validity indicated by strong correlations between RFC and PIQ as well as POI. (though higher than anticipated). Paired t-tests revealed significantly lower PIQ and POI scores compared to the RFC ($p<.001$).

				Correlations on WAIS subtests with CTT1 and CTT2, particularly the CTT1 and the PIQ as well as PSI and the CTT2 and the FSIQ, PIQ, POI, PSI as well as the WMI. Paired t-tests revealed significantly lower FSIQ, VIQ and PIQ scores compared to the CTT's (p<.001).
Shuttleworth-Edwards, A.B., Donnelly, M.J., Reid, I., & Radloff, S.E. (2004). A Cross-cultural study with culture fair normative indications on WAIS-III Digit Symbol—Incidental Learning. Journal of Clinical and Experimental Neuropsychology, 26(7), 921-932. doi 1380-3395/04/2607-921	South African (black African) South African (English)	WAIS-III (Digit- symbol)	**N=68** 40 black African first language 28 English first language African Age range= 19-30 years Education= 15-18 years black African: 20 advantaged education, 20 disadvantaged education English first language: all advantaged education	Results indicated no effect or interaction of language, or education. Normative data was created and compared to WAIS-III norms with no significant differences found.

Demsky, Y.I., Mittenberg, W., Quintar, B., Katell, A.D., & Golden C.J. (1998). Bias in the use of standard American norms with Spanish translations of the Wechsler Memory Scale-Revised. Test. 5(2), 115-121.	Spanish English	Spanish translated WMS-R	**N=100** 50 Spanish-speakers 50 English-speakers Age range = 25-34 years Education = average 11 years	Spanish participants scored significantly lower than the English sample on both verbal and nonverbal subtests (p<.01).
Shuttleworth-Edwards, A.B., Kemp, R.D., Rust, A.L., Muirhead, J.G.L., Hartman, N.P., & Radloff, S.E. (2004b) Cross cultural effects of IQ test performance: a review and preliminary normative indications on WAIS-III test performance. Journal of Clinical and Experimental Neuropsychology. 26: 7 DOI: 10.1080/13803390490510824	South African (black African) South African (English)	WAIS-III	**N=68** Age range= 19-30 years Education black African advantaged: 10= Grade 12 10= Graduate black African disadvantaged: 10= Grade 12 10= Graduate English advantaged: 14= Grade 12 14= Graduate	Low level of education (below graduate level) and poor quality of education impacted on scores on the test for both groups (p<.05). In the black African group, quality of education had more of an impact on scores on the WAIS-III across all subtests (p<.05). Quality of education was seen to be an important factor having more of an impact on scores than language.
Harris, J.G., Cullum, M., & Puente, A.E. (1995). Effects of bilingualism on verbal learning and memory in Hispanic adults, Journal of	American (Spanish) English	Verbal learning lists (English and Spanish)	**N=66** 44= Hispanic bilingual speakers	“Non balanced” individuals recalled significantly less words than the balanced bilingual [F(1,63) = 14.00, p<.001] and

International Neuropsychological Society, 1, 10-16.			(balanced in English and Spanish or non balanced; more proficient in Spanish) 22= monolingual English speakers Age range= 21-50 years Education= 6-16 years	monolingual individuals [F(1,63) = 11.81, p<.001] on the English lists. There were no differences when tested with a list in their dominant language.
Razani, J., Burciaga, J., Madore, M., & Wong, J. (2007). Effects of acculturation on tests of attention and information processing in an ethnically diverse group. Archives of Clinical Neuropsychology, 22, 333-341.	English Hispanic Asian Middle Eastern	Acculturation Rating Scale for Mexican Americans WAIS-III (Digit span and digit symbol) Trail Making Test (TMT) Stroop test Auditory Consonant Trigrams (ACT)	**N=123** 39= monolingual English speakers Age range= 20-72 years Education = 10-16 years 84 = ethnically diverse (Hispanic [43], Asian [20], Middle Eastern [21]) Age range= 18-69 years Education= 9-20 years.	The monolingual group performed better on the Digit Span, [F (1, 122) = 21.45, p < .0001]; Digit Symbol, [F (1, 118) = 4.21, p = .045]; TMT Part B, [F (1,96) = 10.61, p = .002]; Stroop B, [F (1, 119) = 12.55, p = .001]; Stroop C, [F (1, 119) = 5.04, p = .01]; and ACT 18-s delay, [F (1, 119) = 5.82, p = .01]. No differences were found between groups on TMT Part A, [F (1, 107) = 1.36, p = .29]; Stroop A, [F (1, 119) = 2.95,p = .09]; ACT 0-s delay, [F (1, 119) = 0.92, p = .34];

				ACT 3-s delay, [F (1, 119) = .01, p = .99]; or ACT 9-s delay, [F (1,119) = 3.52, p = .06]. Level of acculturation significantly correlated with increased performance on Digit Symbol, Digit Span, TMT Part A, and Stroop B for the ethnically diverse group.
Carstairs, J.R., Myors, B., Shores, E.A., & Fogarty, G. (2006). Influence of language background on tests of cognitive abilities: Australian data, Australian Psychologist, 41(1): 48-54.	Australian (English) Australian (non-English speaking background)	Wechsler Memory Scale-Revised (WMS-R, Wechsler, 1987); Rey Auditory Verbal Learning Test (RAVLT; Lezak, 1996) Wechsler Adult Intelligence Scale-Revised (WAIS-R, Wechsler, 1981) Stockings of Cambridge Test (Cambridge Cognition, 1995) Category Test (Computerised version, Choca, Laatsch, Garside, & Arnemann, 1994) Excluded Letter Fluency Test (Crawford, Wright, & Bate, 1995) Speed and Capacity of Language Processing Test, Form-A (SCOLP; Baddeley, Emslie, & Smith, 1992) Contextual version of the Australian National Adult Reading Test (C-AUSNART; Hennesy, 1995)	**N=116** Non English speaking background: English first language (N=34) English second language (N=42) English speakers (N=40) Age range= 18-34 years Education: Average 13 years	Language background did not influence performance on language and memory functioning tests (WMS-R & RAVLT) or verbal and nonverbal higher order or executive functioning tests (Stockings of Cambridge, Category Test, Excluded letter fluency test). Language background influenced performance on Verbal IQ, Performance IQ and Full scale IQ (WAIS-R) with those from a non-English speaking background with English as a second language scoring the lowest (p<.01). Cultural differences were found to influence Performance IQ rather than language. Language differences were found to

				influence Verbal IQ rather than culture. Language influenced performance on the SCOLP with those from non-English speaking background with English as a second language scoring the lowest (p<.01).
Dugbartey, A.T., Townes, B.D., Mahurin, R.K. (2000). Equivalence of the Colour Trails Test and Trail Making Test in non-native English-speakers. Archives of Clinical Neuropsychology, 15 (5), 425-431.	Turkish (English as a second language)	TMT-A, TMT-B (Army individual test battery, 1994) CTT-1, CTT-2 Stroop colour word test (Golden, 1978) Symbol Search (WISC-III, Wechsler, 1989) Controlled Oral Word Association Test (Benton, 1969) Digit Symbol subtest (WAIS-R, Wechsler, 1981).	**N=64** bilingual Turkish students Age= mean 22.67 years Education= 13-18 years	No significant difference between performance on CTT-1 & TMT-A (Z=-.83, p 5 .40) with a significant correlation (r=.35, p<.01) A significant difference was found between performance on CTT-2 and TMT-B (Z= -4.38, p < .0001), with a moderate correlation (r=.45, p<.01) There was a significant difference between performance on the COWAT and TMT-B (Z= -3.61, p <.001), Results suggest equivalence in task performance for the CTT-1 and TMT-A but not for the CTT-2 and TMT-B.
Tollman, S.G., & Msengana, N.B. (1990). Neuropsychological test: Problems in evaluating the higher mental functioning of Zulu-speaking people using traditional western techniques. South African Journal of Psychology, 20 (1).	South African (Zulu & English speakers)	Luria's Neuropsychological Investigation: LNI (1974)	**N=42** 21 Zulu and English speaking women 21 English speaking women	Length of time in education correlated with higher scores on the LNI.

			Ages and years in education not reported.	
Lu, L. & Bigler, E.D. (2000). Performance on original and a Chinese version of Trail Making Test Part B: A Normative Bilingual Sample, Applied Neuropsychology, 7:4, 243-246. DOI: 10.1207/S15324826AN0704_6.	Chinese English	Trail Making Test (TMT-A and TMT-B) as well as Chinese adaptation of the TMT-B Digit Symbol Coding task (WAIS-III)	**N=60** 30 English speaking students Age range= 21-30 years Education= 15-19 years 30 Chinese speaking students Age range= 24-32 Education= 15-21	No significant difference between groups on TMT-A and Digit Symbol. Significant difference on the TMT-B between groups (p<.05) No significant difference between groups when using the Chinese adapted TMT-B
Lee, T.M.C., Cheung, C.C.Y., Chan, J.K.P., & Chan, C.C.H. (2000). Trail Making across languages, Journal of Clinical and Experimental Neuropsychology, 22(6), 772-778.	Chinese English	TMT-A and B CTT - 1 and 2	**N=84** 52= Chinese-English bilingual 32= English monolingual Age range: 20-50 years. Education: all to university level (16 years)	Significant difference between Chine-English bilingual individuals and English monolingual individuals on TMT-A [F(1,82)=6.61, p=.012] but not TMT-B.

Appendix 2: Semi-structured interview

Semi-structured interview

1. How did you find the testing experience?

2. Did you find the language used in the testing accessible?

- If not, why?

3. Did you feel that you were disadvantaged in any way at any point in the testing?

- Why do you think that you felt disadvantaged?

- Was there something that occurred during testing that led to this feeling?

4. Is there anything that would have made the testing experience easier?

5. If you were to do the test again is there anything that you would have liked to be different?

- why would this have made the experience different?

6. Would you change anything about the instructions?

-Were they easy to follow?

-Did you find anything difficult about the way the questions were asked?

7. Were there any specific parts of the test that caused you to feel anxious? Or were there aspects of the testing situation that led you to feel anxious?

8. Did you worry about any of the tasks?

- What did you worry about? (e.g. getting the answers right, understanding of English, understanding of why you were doing the test, understanding of the situation)

9. Was there anything about doing this test that surprised you?

Appendix 3: Screening Questionnaire

Participant Questionnaire
Please ask if you need any help completing any of the questions below.

Participant ID: ____________
Gender: Male Female
Handedness: Left Right
Age: 18-32 33-50 51-64 65-74

Education: (please tick)
__ Less than secondary/high school education
__ Secondary education completed (i.e. O-Levels, GCSE
__ College or vocational training/apprenticeship (i.e. lasting 6 months or more)
__ University degree or higher

Employment: (please tick)
__ University degree required
__ Skilled (apprenticeship, college, 6+ months training required)
__ Unskilled (short training)
__ Unemployed (for past 6+ months - if less use above categories)
__ On disability benefit (for past 6+ months - if less use above categories)
__ Student
__ At home parent

Please tick Full-time__ Part-time__ N/A__

Health Questionnaire: Please tick all that apply.

We are asking about your health status as these are factors that can sometimes have an impact on cognitive performance.

__ Attention Deficit Disorder
__ Autistic Spectrum/Aspergers
__ Brain injury
__ Dementia
__ Dependency with alcohol/drugs
__ Depression
__ Diabetes
__ Epilepsy
__ Hearing impairment (Do you have a hearing aid? **Y** **N**)
__ Hypertension
__ Loss of consciousness following a knock to the head?
__ Neurological condition (MS, motor neurone disease, Parkinson’s disease)
__ Psychiatric condition
__ Smoker (How many per day_____)

__ Stroke
__ TIA
__ Viral infection (e.g. meningitis)
__ Visual impairment (Do you wear glasses or contact lenses **Y** **N**)

Appendix 4: Credibility Check using Kappa Measure of Agreement

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Rater 1 * Rater 2	18	100.0%	0	0.0%	18	100.0%

Rater 1 * Rater 2 Crosstabulation

			Rater 2		
			no	yes	Total
Rater 1	no	Count	7	0	7
		% within Rater 1	100.0%	0.0%	100.0%
	yes	Count	0	11	11
		% within Rater 1	0.0%	100.0%	100.0%
Total		Count	7	11	18
		% within Rater 1	38.9%	61.1%	100.0%

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	1.000	.000	4.243	.000
N of Valid Cases		18			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Appendix 5: Ethics Approval

This has been removed from the electronic copy

Appendix 6: Consent Form

INFORMED CONSENT FORM

Exploration of the impact of culture, language and testing experience on a Neuropsychological Test

I confirm that (please tick box as appropriate):

1.	I have read and understood the information about the project, as provided in the Information Sheet	□
2.	I have been given the opportunity to ask questions about the project and my participation.	□
3.	I voluntarily agree to participate in the project.	□
4.	I understand I can withdraw at any time without giving reasons and that I will not be penalised for withdrawing nor will I be questioned on why I have withdrawn.	□
5.	The procedures regarding confidentiality have been clearly explained (e.g. use of names, pseudonyms, anonymisation of data, etc.) to me.	□
6.	If applicable, separate terms of consent for interviews, audio, video or other forms of data collection, including the use of quotations in the final report have been explained and provided to me.	□
7.	The use of the data in research, publications, sharing and archiving has been explained to me.	□
8.	I understand that other researchers will have access to this data only if they agree to preserve the confidentiality of the data and if they agree to the terms I have specified in this form.	□

9.	I am aware that participation in this study involves completion of some tests which are routinely used as screens for clinical conditions. I understand that these tests are not used for diagnostic purposes in this study. I also understand that the researchers cannot inform participants of individual test scores, but I can discuss any concerns with the researcher at the end of the study	☐
10.	I, along with the Researcher, agree to sign and date this informed consent form.	☐

____

Participant's Name (Printed)*

______________________________ _________________________

Participant's signature* Date

______________________________ _________________________________

Name of person obtaining consent (Printed) Signature of person obtaining consent

*Participants wishing to preserve some degree of anonymity may use their initials (from the British Psychological Society Guidelines for Minimal Standards of Ethical Approval in Psychological Research)

Appendix 7: Information sheet

Canterbury Christ Church University

PARTICIPANT INFORMATION SHEET

Exploration of the Impact of Culture, Language and Testing Experience on a Neuropsychological Test

INVITATION

You are being asked to take part in a research study aiming to test people whose first language is not English on a new psychological test. The research aims to see if there are differences between scores for people whose first language is English and those with English as a second language. We would also like to ask you about your experiences of completing the test and if there were any difficulties that may relate to language or understanding.

This research is part of my doctoral training in clinical psychology at Salomons, Canterbury Christ Church University. I am supervised by Dr Jerry Burgess (Canterbury Christ Church University) and Dr David Eley (Kent & Medway NHS Partnership Trust) who are both Clinical Psychologists.

This research project has received ethical approval from Salomons, Canterbury Christ Church University ethics panel.

WHO HAVE WE ASKED TO PARTICIPATE?

We have invited people aged 18-65 whose first language and culture is not English.

WHO MUST WE EXCLUDE?

People who have had a head or brain injury in the past will be unable to take part in this study.

People who have English as a first language will be unable to take part in this study.

WHAT WILL HAPPEN?

In this study, you will be asked to complete a series of psychological tasks. You will be asked a series of questions that look at different areas of functioning for example memory and language. The tests will take approximately 1 hour to complete. It is preferable for the tests to be completed all in one go, however if you need a break please let me know.

Language, Culture and Neuropsychological Tests

Psychological tests give an indication of current functioning and tests such as the one we will be doing today are often used to help us to understand a person's functioning after they have sustained an injury to the head. In developing such tests we need to test people who have not experienced any head injuries or had a stroke to allow us to see how people perform differently after a head injury.

Psychological tests give an indication of how a person performs on a given day at a given time. It is a snapshot of performance and can be affected by things such as tiredness, hunger, stress or anxiety. It is important to remember that it is just a snapshot and on a different day at a different time you may perform better or worse.

After we have completed the test I would like to ask you a series of questions about your experience of taking the test. This should take approximately 15 minutes. At the end of the test I will ask you if you would be happy to take part in the interview. Interviews will be audio taped and transcribed.

HOW LONG WILL IT TAKE?

The research will take place somewhere convenient for you and will typically take 1 ½ to 2 hours.

YOUR RIGHTS

You may decide to stop being a part of the research study at any time without explanation. You have the right to ask that any data you have supplied to that point be withdrawn/destroyed.

You have the right to omit or refuse to answer or respond to any question that is asked of you.

You have the right to have your questions about the procedures answered (unless answering these questions would interfere with the study's outcome). If you have any questions as a result of reading this information sheet, please ask before the test begins.

BENEFITS AND RISKS

Participation in this study involves completion of tests which are used as preliminary screens for clinical conditions involving cognitive impairment. Scores from these tests are not used for diagnostic purposes in this study. Though it is not possible to provide feedback of individual scores to participants, you will have the opportunity to discuss any concerns that you may have after completing the test.

CONFIDENTIALITY/ANONYMITY

The data we collect do not contain any personal information about you except your age, language and country of origin and number of years in education. All test sheets will be numbered with a code.

Interview recordings will not include any identifiable information and will be stored on an encrypted memory stick.

All research material will be stored at the university for 10 years following the completion of the project, after which it will be destroyed.

Data will be recorded in the writing up of the study and may be presented within journals or presented at conferences. Quotations from interviews may be used in the final report however, no personal information will be included in write-ups or subsequent presentations about the study.

If you would like to receive a copy of a summary of the results of the research you can leave your email or home address with the researcher. This personal data will be stored on a separate database and will not be connected to the research data.

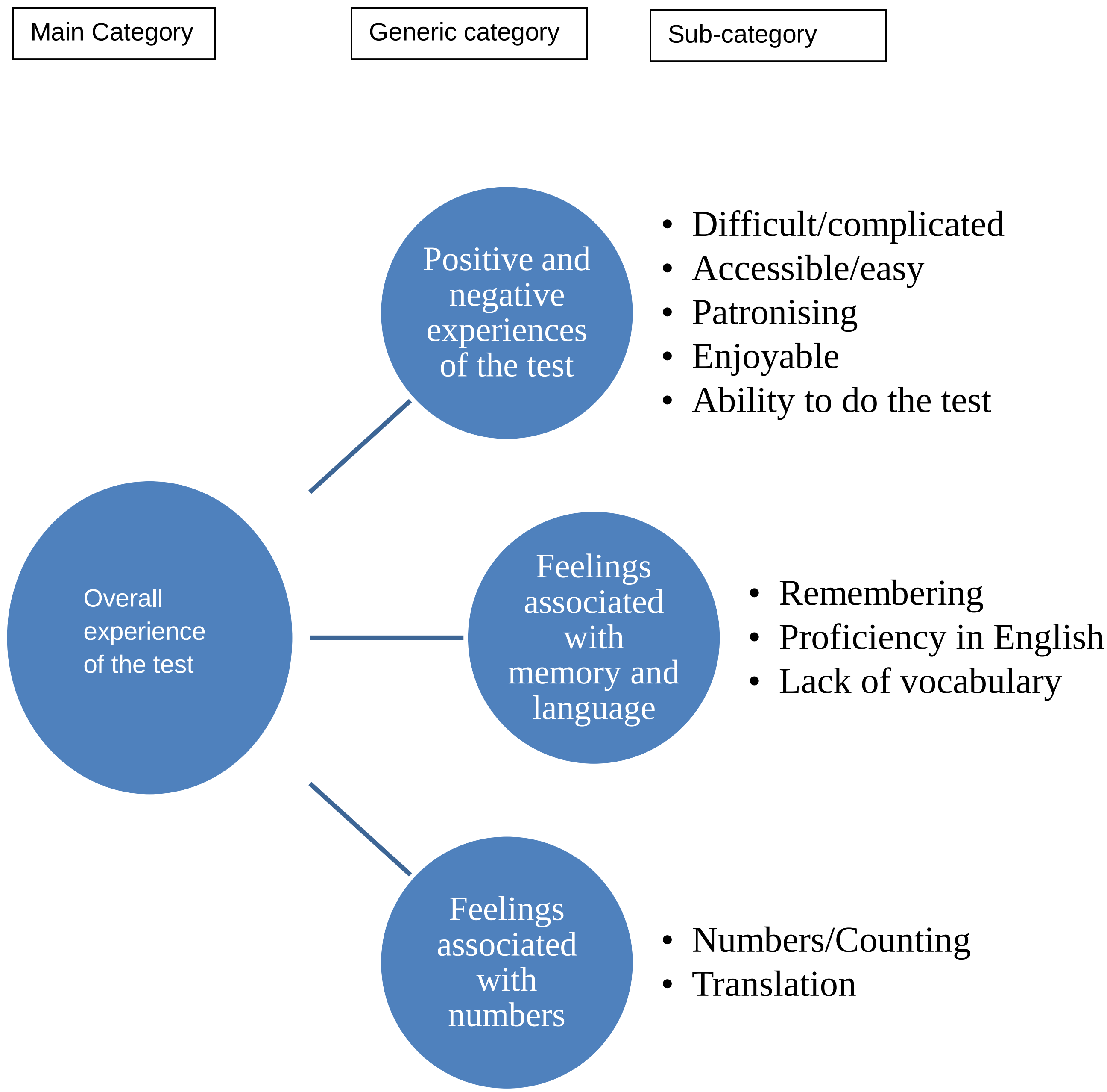
FOR FURTHER INFORMATION

I will be glad to answer your questions about this study at any time. You may contact me at c.haddlesey505@canterbury.ac.uk or by post at Salomons Centre for Applied Psychology, Canterbury Christ Church University, Runcie Court, David Salomons Estate, Broomhill Road, Tunbridge Wells, TN3 0TF.

Appendix 8: Language History Questionnaire

This has been removed from the electronic copy

Appendix 9: Diagrammatic abstraction process



Appendix 10: Categories and sample quotes

Generic Category	Sub-category (number of participants reporting)	Sample Quotes
Positive and negative experiences of the test	Difficult/complicated (7)	"it was fine, a bit complicated some of them" "…were more difficult than expected"
	Accessible/Easy (14)	"definitely yeah, it was very easy"
	Patronising (3)	"sometimes a little bit insulting, I suppose when you think about rather simple questions and then see how you perform"
	Enjoyable (9)	"umm…it was fun" "I liked it"
	ability to do the test (10)	"R: was the worry about getting the answers right? L: yes"
Feelings associated with memory and language	Remembering (10)	"oh god, I'm not going to remember this" "how bad my memory is" "I was quite nervous to start with, I couldn't work out if I got dementia"

	proficiency in English (6)	"…which is really odd, because I consider myself you know, having pretty good English"
	lack of vocabulary (4)	"I mean apart from my lack of vocabulary for, what was that tool again, sadler? Was it?"
Feelings associated with numbers	Numbers/Counting (11)	"I just saw numbers and I was like oh no" "umm number counting and remembering words, is kind of an issue…"
	Translation (2)	"…but it's something that could…because of the translating…as well as trying to remember

Appendix 11: Sample transcript with codes

This has been removed from the electronic version

Appendix 12: Neuropsychological Rehabilitation author guidelines

Instructions for authors

Thank you for choosing to submit your paper to us. These instructions will ensure we have everything required so your paper can move through peer review, production and publication smoothly. Please take the time to read them and follow the instructions as closely as possible.

AUTHORSERVICES
Supporting Taylor & Francis authors

Should you have any queries, please visit our Author Services website or contact us atauthorqueries@tandf.co.uk.

SCHOLARONE MANUSCRIPTS™

This journal uses ScholarOne Manuscripts (previously Manuscript Central) to peer review manuscript submissions. Please read the guide for ScholarOne authors before making a submission. Complete guidelines for preparing and submitting your manuscript to this journal are provided below.

Use these instructions if you are preparing a manuscript to submit to *Neuropsychological Rehabilitation*. To explore our journals portfolio, visit http://www.tandfonline.com/, and for more author resources, visit ourAuthor Services website.

Neuropsychological Rehabilitation considers all manuscripts on the strict condition that

- the manuscript is your own original work, and does not duplicate any other previously published work, including your own previously published work.
- the manuscript has been submitted only to *Neuropsychological Rehabilitation*; it is not under consideration or peer review or accepted for publication or in press or published elsewhere.
- the manuscript contains nothing that is abusive, defamatory, libellous, obscene, fraudulent, or illegal.

Please note that *Neuropsychological Rehabilitation* uses CrossCheck™ software to screen manuscripts for unoriginal material. By submitting your manuscript to *Neuropsychological Rehabilitation* you are agreeing to any necessary originality checks your manuscript may have to undergo during the peer-review and production processes.

Any author who fails to adhere to the above conditions will be charged with costs which *Neuropsychological Rehabilitation* incurs for their manuscript at the discretion of *Neuropsychological Rehabilitation*'s Editors and Taylor & Francis, and their manuscript will be rejected.

This journal is compliant with the Research Councils UK OA policy. Please see the licence options and embargo periods here.

Contents List

Manuscript preparation

1. Journal-specific guidelines

- This journal accepts original (regular) articles, scholarly reviews, and book reviews.
- The style and format of the typescripts should conform to the specifications given in the Publication Manual of the American Psychological Association (6th ed.).
- There is no word limit for manuscripts submitted to this journal. Authors should include a word count with their manuscript.

2. General guidelines

↑Back to top.

- Manuscripts are accepted in English. Oxford English Dictionary spelling and punctuation are preferred. Please use double quotation marks, except where "a quotation is 'within' a quotation". Long quotations of words or more should be indented without quotation marks.
- Manuscripts should be compiled in the following order: title page; abstract; keywords; main text; acknowledgements; references; appendices (as appropriate); table(s) with caption(s) (on individual pages); figure caption(s) (as a list).
- Abstracts of 150-200 words are required for all manuscripts submitted.
- Each manuscript should have up to 5 keywords.
- Search engine optimization (SEO) is a means of making your article more visible to anyone who might be looking for it. Please consult our guidance here.
- Section headings should be concise.
- All authors of a manuscript should include their full names, affiliations, postal addresses, telephone numbers and email addresses on the cover page of the manuscript. One author should be identified as the corresponding author. Please give the affiliation where the research was conducted. If any of the named co-authors moves affiliation during the peer review process, the new affiliation can be given as a footnote. Please note that no changes to affiliation can be made after the manuscript is accepted. Please note that the email address of the corresponding author will normally be displayed in the article PDF (depending on the journal style) and the online article.
- All persons who have a reasonable claim to authorship must be named in the manuscript as co-authors; the corresponding author must be authorized by all co-authors to act as an agent on their behalf in all matters pertaining to publication of the manuscript, and the order of names should be agreed by all authors.
- Biographical notes on contributors are not required for this journal.
- Please supply all details required by any funding and grant-awarding bodies as an Acknowledgement on the title page of the manuscript, in a separate paragraph, as follows:
  - *For single agency grants:* "This work was supported by the [Funding Agency] under Grant [number xxxx]."

  - *For multiple agency grants:* "This work was supported by the [Funding Agency 1] under Grant [number xxxx]; [Funding Agency 2] under Grant [number xxxx]; and [Funding Agency 3] under Grant [number xxxx]."
- Authors must also incorporate a Disclosure Statement which will acknowledge any financial interest or benefit they have arising from the direct applications of their research.
- For all manuscripts non-discriminatory language is mandatory. Sexist or racist terms must not be used.
- Authors must adhere to SI units. Units are not italicised.
- When using a word which is or is asserted to be a proprietary term or trade mark, authors must use the symbol ® or TM.

2. Style guidelines

↑Back to top.

- Description of the Journal's reference style.
- Guide to using mathematical scripts and equations.
- Word templates are available for this journal. If you are not able to use the template via the links or if you have any other template queries, please contact authortemplate@tandf.co.uk.
- Authors must not embed equations or image files within their manuscript

3. Figures

↑Back to top.

- Please provide the highest quality figure format possible. Please be sure that all imported scanned material is scanned at the appropriate resolution: 1200 dpi for line art, 600 dpi for grayscale and 300 dpi for colour.
- Figures must be saved separate to text. Please do not embed figures in the manuscript file.
- Files should be saved as one of the following formats: TIFF (tagged image file format), PostScript or EPS (encapsulated PostScript), and should contain all the necessary font information and the source file of the application (e.g. CorelDraw/Mac, CorelDraw/PC).
- All figures must be numbered in the order in which they appear in the manuscript (e.g. Figure 1, Figure 2). In multi-part figures, each part should be labelled (e.g. Figure 1(a), Figure 1(b)).
- Figure captions must be saved separately, as part of the file containing the complete text of the manuscript, and numbered correspondingly.
- The filename for a graphic should be descriptive of the graphic, e.g. Figure1, Figure2a.

4. Publication charges

↑Back to top.

Submission fee

There is no submission fee for *Neuropsychological Rehabilitation*.

Page charges

There are no page charges for *Neuropsychological Rehabilitation*.

Colour charges

Colour figures will be reproduced in colour in the online edition of the journal free of charge. If it is necessary for the figures to be reproduced in colour in the print version, a charge will apply. Charges for colour figures in print are £250 per figure ($395 US Dollars; $385 Australian Dollars; 315 Euros). For more than 4 colour figures, figures 5 and above will be charged at £50 per figure ($80 US Dollars; $75 Australian Dollars; 63 Euros).

Depending on your location, these charges may be subject to Value Added Tax.

5. Reproduction of copyright material

↑Back to top.

If you wish to include any material in your manuscript in which you do not hold copyright, you must obtain written permission from the copyright owner, prior to submission. Such material may be in the form of text, data, table, illustration, photograph, line drawing, audio clip, video clip, film still, and screenshot, and any supplemental material you propose to include. This applies to direct (verbatim or facsimile) reproduction as well as "derivative reproduction" (where you have created a new figure or table which derives substantially from a copyrighted source).

You must ensure appropriate acknowledgement is given to the permission granted to you for reuse by the copyright holder in each figure or table caption. You are solely responsible for any fees which the copyright holder may charge for reuse.

The reproduction of short extracts of text, excluding poetry and song lyrics, for the purposes of criticism may be possible without formal permission on the basis that the quotation is reproduced accurately and full attribution is given.

For further information and FAQs on the reproduction of copyright material, please consult our Guide.

6. Supplemental online material

↑Back to top.

Authors are encouraged to submit animations, movie files, sound files or any additional information for online publication.

- Information about supplemental online material

Manuscript submission

↑Back to top.

All submissions should be made online at the *Neuropsychological Rehabilitation* Scholar One Manuscriptswebsite. New users should first create an account. Once logged on to the site, submissions should be made via the Author Centre. Online user guides and access to a helpdesk are available on this website.

Manuscripts may be submitted in any standard editable format, including Word and EndNote. These files will be automatically converted into a PDF file for the review process. LaTeX files should be converted to PDF prior to submission because ScholarOne Manuscripts is not able to convert LaTeX files into PDFs directly. All LaTeX source files should be uploaded alongside the PDF.

Click here for information regarding anonymous peer review.

Copyright and authors' rights

↑Back to top.

To assure the integrity, dissemination, and protection against copyright infringement of published articles, you will be asked to assign us, via a Publishing Agreement, the copyright in your article. Your Article is defined as the final, definitive, and citable Version of Record, and includes: (a) the accepted manuscript in its final form, including the abstract, text, bibliography, and all accompanying tables, illustrations, data; and (b) any supplemental material hosted by Taylor & Francis. Our Publishing Agreement with you will constitute the entire agreement and the sole understanding between you and us; no amendment, addendum, or other communication will be taken into account when interpreting your and our rights and obligations under this Agreement.

Copyright policy is explained in detail here.

Free article access

↑Back to top.

As an author, you will receive free access to your article on Taylor & Francis Online. You will be given access to the *My authored works* section of Taylor & Francis Online, which shows you all your published articles. You can easily view, read, and download your published articles from there. In addition, if someone has cited your article, you will be able to see this information. We are committed to promoting and increasing the visibility of your article and have provided guidance on how you can help. Also within *My authored works*, author eprints allow you as an author to quickly and easily give anyone free access to the electronic version of your article so that your friends and contacts can read and download your published article for free. This applies to all authors (not just the corresponding author).

Reprints and journal copies

↑Back to top.

Corresponding authors receive a complimentary copy of the issue containing their article. Complimentary reprints are available through Rightslink® and additional reprints can be ordered through Rightslink® when proofs are received. If you have any queries about reprints, please contact the Taylor & Francis Author Services team at reprints@tandf.co.uk. To order a copy of the issue containing your article, please contact our Customer Services team at Adhoc@tandf.co.uk

Open Access

↑Back to top.

Taylor & Francis Open Select provides authors or their research sponsors and funders with the option of paying a publishing fee and thereby making an article permanently available for free online access – *open access* – immediately on publication to anyone, anywhere, at any time. This option is made available once an article has been accepted in peer review.

Appendix 13: Final Report on the Research for Ethics board

Dear Ethics Committee Chair,

I am writing to you following completion of my project exploring the impact of culture and language on a neuropsychological test (the Short Parallel Assessments of Neuropsychological Status: SPANS).

Enclosed is a copy of the summary of the project for participants. This provides an overview of the methodology and results of the project. The project aimed to explore whether there were differences between participants who spoke English as a first language and participants who spoke English as a second language on the indices of the SPANS. At the outset of the project participants with a first language of a non-western, South Asian origin were to be recruited, however due to difficulties recruiting within this group, the recruitment was extended to participants whose first language was not English and ultimately all the participants were of a Western European origin. The enclosed summary is to be sent to all participants who indicated interest in the results of the study. The research will be prepared for submission to Neuropsychological Rehabilitation for publication.

Yours sincerely,

Claire Haddlesey

Enc: Summary of project for participants

Appendix 14: Summary of the study for participants

Dear …,

Thank-you for taking part in my study exploring the impact of language and culture on a psychology test, the SPANS. As part of the study I met with you, and 22 other individuals who first language was not English. You were asked to complete a language history questionnaire to give an overview of the length of time you had been speaking English, where you completed your early education and your perceived ability in different areas of speaking English. You completed an assessment of general ability that has been researched to be valid with different populations as well as the SPANS.

The SPANS has been developed for use with people who have experienced a neurological disorder such as a stroke or brain injury. We wanted to test whether there would be any parts of the test that may be difficult if someone's first language was not English. The results of the tests completed with you and the other people whose first language was not English was compared with data for native-English-speaking people. This showed us that there were differences only on the part of the SPANS which looks at language ability and more specifically the Naming and Repetition subtests. The Naming subtest involved you naming six objects and the Repetition subtest involved you repeating back different phrases. These two tasks were seen to be slightly more difficult when English was not a first language. A third task which showed differences between the two groups was the Writing Sentences task, which was scored higher by the English second language group. These findings suggest that generally the SPANS is a fair task to use with people whose first language is not English, if consideration is taken to difficulties that may present with the Naming and Repetition subtests.

The educational levels of the English second language were higher than the English first language group which may have influenced the results. For people whose educational levels are lower, there may be additional considerations necessary and further research into this is necessary.

Some of the English second language group completed an interview about their experiences of completing the test in a second language. The analysis of these interviews revealed that overall experiences of the test were positive and most found the test enjoyable and that the language used in the testing was clear. Some people found aspects of the testing situation worrying or anxiety provoking particularly in relation to remembering information or related to numbers. These are things for people using the SPANS in the future to be aware of when testing a person in their second language.

Once again I would like to thank-you for your valuable time and participation in this study.

Yours sincerely,

Claire Haddlesey

Appendix 15: Extracts from Research Diary

February 2014

Found out the Jerry Burgess has recently published a new neuropsychological assessment and is interested in completing work exploring the use of the test with individuals from different language backgrounds. Approached him to enquire as to whether I could be involved for my MRP. Sparked interest from when completing neuro assessments with translators and often told to interpret test results with caution – felt a need to explore the validity of tests with people whose first language is not English further.

Initial meeting with Jerry to discuss my interest in the project as well as feasibility. Discussions about whether to use a single language group within a non-western culture or more broadly those with English as a second language.

March 2014

Meeting with Jerry to run through the administration of the SPANS. Practice run through with Jerry offering assistance in terms of accurate procedure and scoring considerations.

June 2014

MRP Proposal review at Salomons. Reviewers expressed interest in the project but suggested using an additional measure that has been evidenced as minimally impacted upon by culture to give an indication of L2 group's level of cognitive functioning. Discussions with Jerry and Dave about this and a decision made to use the Raven's Standard Progressive Matrices.

November 2014 – January 2015

Ethics submitted and approved with minor conditions. Ethics process felt much smoother than the review process, which actually may not have been the case without the review.

Initial scoping for participants involved contact through a relative, with a sample of teachers from a South Asian background. After initial contact it was revealed that all had English as a first language. Contact made with two universities in London who indicated that they had no students with a first language of a South Asian background.

January 2015-June 2015

Contact made with numerous community and religious organisations in London regarding recruitment, with no responses. Advertisements placed in local libraries and further universities contacted. One university placed an advert in their newsletter and two people were in contact however after one email to them both, had no further response. Beginning to panic at this stage that I am getting nowhere with recruitment and that maybe this language group is not accessible.

July 2015

Discussions with Dave and Jerry about recruitment. Suggestion from Dave that I widen the search to anyone with a second language that is English. Dave had thought that he might be able to recruit staff from a hospital he had worked in previously, but again there was no uptake on the project. This idea was taken to Jerry who agreed it may be beneficial to widen the language criteria for participants.

August 2015 – March 2016

Participants with second language as English recruited through word of mouth and contact with local universities and libraries. Word of mouth proved to be the most successful method of recruitment. A local language school was contacted and they were confident that a number of their higher level English students would be able to take part, however again contact ceased with me and no participants were suggested.

September 2015

Missed the formative deadline for section A draft. This is due to problems with recruitment but I am still annoyed with myself as missing deadlines is not like me. Feeling the pressure of doing this level of project alongside clinical work and other demands of the course.

January 2016

First draft section A finally completed. Received prompt feedback which was very useful. Skype conversation with Jerry about how to reduce the word count by either tightening exclusion criteria or making synthesis of the data more succinct.

March 2016

Data collection completed without as many participants as expected. Very disappointed to not get the desired numbers for decent power but time is running short and there is little time now to get everything analysed and written-up.

Interesting interview data – initial read-throughs suggest that a lot of those interviewed worry about numbers, not as much language and that digit span tasks are more complicated due to double translations. Richness of interview data does not feel to be there for thematic analysis, a lot of responses seem more quantitative in nature – qualitative content analysis feels more appropriate in this case.