Jennifer Attwood BSc (Hons)

ASSESSING FOR COGNITIVE IMPAIRMENT IN PEOPLE WITH AN ACQUIRED BRAIN INJURY: VALIDATION OF A BRIEF NEUROPSYCHOLOGICAL ASSESSMENT BATTERY

Section A: Assessment of cognitive impairment after acquired brain injury: A review of existing brief comprehensive measures
5466 words

Section B: Reliability, validity and factor structure of the Short Parallel Assessments of Neuropsychological Status (SPANS)
7493 words

Section C: Critical appraisal of the research process
1978 words

Overall Word Count: 14,907

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

JULY 2013

SALOMONS
CANTERBURY CHRIST CHURCH UNIVERSITY

Acknowledgements

I would like to thank all the people who have participated in this study over the last several years and without whom the development of the SPANS would not have been possible.

I would also like to thank a number of people who have supported me throughout this research process. My internal supervisor: Dr Jerry Burgess, who has worked incredibly hard and diligently over the last several years in conceiving, designing, developing, trialling and validating the SPANS and who has given me the benefit of his knowledge and expertise – not to mention promptly answering my millions of questions! To my external supervisor: Dr Seb Potter, for his advice and support prior to and throughout the project and for consistently wearing the "Viva Fez". To Dr Sabina Hulbert for her crucial statistics consultancy; patience and enthusiasm in helping me understand new statistical concepts; and calming presence when I was pulling my hair out!

Lastly, I would like to thank my manager, Dr Fergal Jones; my family and friends who have support me throughout the three years of clinical training.

Summary of MRP

Section A identifies and critically evaluates existing tools for assessing cognitive impairment in people who have experienced an acquired brain injury. The clinical utility and psychometric properties of each measure is discussed and recommendations are made for the need for further validation of these measures or the development of a new comprehensive assessment tool.

Section B is a psychometric validation study of a new neuropsychological assessment: the SPANS. The internal reliability, discriminative validity, classification accuracy, and factor structure of the measure was assessed.

Section C provides a critical and reflective account of the research process and the authors own learning. Directions for future research and clinical implications of the study are also considered.

Contents

Section D: List of tables and figure

Section E: Appendices

Section A:

Assessment of cognitive impairment after acquired brain injury: A review of existing brief comprehensive measures

Jennifer Attwood

July 2013

Word count: 5466

A thesis submitted in partial fulfilment of the requirements of

Canterbury Christ Church University for the degree of

Doctor of Clinical Psychology Abstract

Abstract

Assessment of cognitive functioning following an acquired brain injury is an important aspect of neuro-rehabilitation services and can help guide treatment planning. Due to resource limitations in services and patients' difficulty tolerating lengthy assessments due to factors such as fatigue, there is a need for brief but comprehensive measures of cognitive function that can be easily administered by a range of health professionals. The current review aimed to identify and critically evaluate the psychometric properties of existing brief but comprehensive measures of cognitive function that have been validated in an acquired brain injury population. The literature search identified 15 papers covering four different tests: Addenbrooke's Cognitive Examination – Revised (ACE-R); the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS); the Cognistat; and the Neuropsychological Assessment Battery - Screening Module (NAB-SM). The review found that there is some evidence to support the use of the RBANS or NAB-SM with people with an acquired brain injury, though these measures have limitations in terms of construct validity and reliability of indices respectively. Further research is required to increase the evidence base for the use of either of these measures with people with an acquired brain injury. Alternatively, a new purpose-designed tool could be developed that aims to address the limitations of existing measures.

Keywords: Acquired brain injury, neuropsychological assessment, psychometrics, cognitive impairment, screening tools

Introduction

Acquired brain injury:

Acquired brain injury (ABI) refers to any injury to the brain that has occurred since birth (i.e. is not a neurodevelopmental disorder) and is non-degenerative (Headway: the brain injury association). It includes traumatic brain injuries (TBI), such as those resulting from a road traffic accident or assault, and non-traumatic injuries, for example cerebrovascular accidents (stroke). According to the World Health Organisation [WHO], (2006) neurological disorders contribute 6.3% to the global burden of disease. The second most common cause of death worldwide is accounted for by strokes (Sarti, Rastenyte, Cepaitis, & Tuomilehto, 2000), whilst TBIs are responsible for the majority of deaths and disability in children and young adults (WHO, 2006). In the UK alone, it is estimated that there are approximately 950,000 working-age adults living with on-going impairments as a result of TBI or stroke (Headway: the Brain Injury Association).

Cognitive impairment:

Cognitive complaints are common following an ABI. Approximately 60% of people who have suffered a stroke will experience some level of cognitive dysfunction (Jin, Legge, Ostbye, Feightner, & Hachinski, 2006). Research has demonstrated a linear relationship between injury severity and extent and persistence of cognitive dysfunction (Dikmen, Machamer, Winn, & Temkin, 1995). Impairments may include confusion, disorientation, attention deficits, memory problems, language and communication difficulties, impaired judgement, poor organisation and planning, and inflexibility of thinking (Wallesch, Curio, Galazky, Jost, & Synowitz, 2001). Early cognitive status is predictive of overall outcome and

can determine the need for further rehabilitation or care post-discharge (Dikmen, McLean, Temkin, & Wyler, 1986). Greater cognitive impairment is associated with the risk of poor return to work (Shames, Treger, Ring, & Giaquinto, 2007) and limited community integration (Kneipp & Rubin, 2007).

Neuropsychological assessment:

Neuropsychological assessment is "the normatively informed application of performance-based assessments of various cognitive skills" (Harvey, 2012, p. 91). It has been shown to affect the management of patients and improve outcomes (Chelune, 2010). Due to individual differences in functional neuroanatomy (Heilman & Valenstein, 2003), assessment is essential to describe the cognitive-behavioural expression of an injury for that individual (Schoenberg, 2011). As well as providing information on how a person might cope with medication regimes (Hinkin et al., 2002), driving (Schanke & Sundet, 2000), or self-care (McCue, Rogers, & Goldstein, 1990); neuropsychological assessments have also been shown to be predictive of employment outcome (Ponsford et al., 2008), likelihood of seizures (Sawrie et al., 1998), and risk of further cognitive decline (Herrmann, Goodwin, & Ebmeier, 2007).

Psychometric theory

For neuropsychological assessments to be meaningful, the knowledge they produce must be reliable and valid. To prove that this is the case, tests must be validated using psychometric methods (Anastasi & Urbina, 1997; Russell, Russell, & Hill, 2005). Reliability means that any change in scores is due to change in the construct being measured, as opposed to random

error or influence of confounding variables. Reliability is not an absolute property of tests but one of degree that must be assessed with different populations in different contexts (Schoenberg, 2011). Validity refers to the extent that the test measures what it claims to. It is a property not of the test itself, but the conclusions drawn from the test (Schoenberg, 2011). Evidence for validity is generally considered within a tripartite model consisting of: content; construct (convergent/divergent); and criterion-related (predictive/concurrent) evidence (Schoenberg, 2011; Anastasia & Urbina, 1997). An example of criterion-related evidence is classification accuracy statistics (or diagnostic validity – these are described in more detail in the literature review).

Another principal of neuropsychological assessment is normative comparison (Harvey, 2012). Clinicians need to know how a comparable person would be expected to score on a test in order to infer whether a patient's performance is outside normal limits. If assessing multiple domains, clinicians need to know what degree of variability is normal. If normative data are not representative of the patient being assessed, interpretation may be invalid. Similarly, if tests have not been normed on a single sample (as is the case in flexible batteries that pool different tests), comparing test scores is invalid (Harvey, 2012). This is one of the key reasons why Russell et al., (2005) argue that flexible batteries undermine the fundamental principles of neuropsychological assessment and conclude that fixed and co-normed batteries should be employed.

Clinical practice:

In reality it is not practicable for every patient to undergo a full neuropsychological assessment. Issues of fatigue and attention deficits make participating in lengthy testing problematic for patients (McKay, Wertheimer, Fichtenbery, & Casey, 2008). Rapid changes in the early stages following an ABI may render cognitive assessments obsolete, whilst assessing towards the end of admission does little to inform treatment planning (Nabors, Millis, & Rosenthal, 1997). Furthermore, resource issues have resulted in a trend for decreasing lengths of stay in hospitals and rehabilitation settings (Nabors et al., 1997), as well as limiting access to neuropsychological expertise (McMillan & Ledder, 2001). These pressures have resulted in a need for neuropsychological measures that can identify the presence of cognitive impairment and the specific domains affected (Lezak, 2004). For clinical utility, the tool needs to be quick and easy to administer by any health care professional, be repeatable in order to document change, and have the potential to guide recommendations for rehabilitation (Nabors et al., 1997).

Rationale for review:

There is a need for brief but comprehensive and psychometrically sound neuropsychological assessments that can be used to determine cognitive function in patients who have had an ABI. Though reviews have been conducted about screening tools for dementia (Cullen, O'Neill, Evans, Coen, & Lawlor, 2007), neurobehavioural disability (Woods, Alderman, & Williams, 2008) and executive function (Poulin, Korner-Bitensky, & Dawson, 2013), currently there does not exist any review of tools for ABI. This current review therefore aims to search the literature to identify and critically evaluate the psychometric properties of neuropsychological assessments which have been researched with an ABI population.

Methods:

A literature search was conducted which resulted in 15 papers to be included in this review (see Appendix A). Neuropsychological tests (NTs) were included if they: 1) were designed to assess for cognitive impairment in multiple domains and provide separate scores for each domain; 2) were performance-based measures given direct to patients; 3) could be administered in one hour or less; and 4) had been published in English. Individual standardised tests, flexible batteries, standardised batteries with administration times of over one hour, tests which only give a global score, or unpublished tests were excluded.

Papers relating to each measure were included if: a) the main aim of the study was to investigate the psychometric properties of the test; and b) participants were predominantly 16 – 65 years old, English-speaking, with an ABI. Studies including older adults were retained if the mean age was less than 65 years old. Studies conducted solely with children and adolescents or older adults were excluded. Studies with neurodevelopmental or neurodegenerative conditions or physical illnesses with associated cognitive impairments (e.g. HIV) were also excluded.

Literature search results:

Four tests discussed in 15 papers were identified which met the selection criteria: Addenbrooke's Cognitive Examination - Revised (ACE-R: Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006); the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS: Randolph, 1998); the Cognistat (formerly Neurobehavioural Cognitive Status Examination: Kiernan, Mueller, Langston, & Van Dyke, 1987); and the

Neuropsychological Assessment Battery - Screening Module (NAB-SM: Stern & White, 2003). For each test, a brief description of the design and development is provided[1] followed by a critical evaluation (based on guidelines by Fritz & Wainner, 2001; Greenhalgh, 1997; and Jaeschke, Guyatt, & Sackett, 1994) of the psychometric research with ABI populations. Key information for each test can be seen in Table 1 for ease of comparison (see Appendix B for a more detailed summary).

[1] Where this information was not available through the retrieved articles a supplementary PsychINFO search was conducted using the individual test name in order to identify the original validation paper.

Table 1: Summary of key information for each test

	ACE-R	RBANS	NAB-SM	Cognistat
Designed for:	Dementia screen	Dementia screen	Generic cognitive screen. Determine need for full NAB	Generic cognitive screen.
Normative data:	N = 63, aged 50 -75, 12.7 years education (not age/education scaled)	N = 540, aged 20-89; age-scaled scores and information on education effects provided	N= 1,448; age, gender and education scaled-scores provided	N = 116 adults, aged 20 to 92
Admin time:	Mean 15 mins	20-30 mins	24-45 mins	20-30 mins
Constructs measured:	A/O; M; VF; L; V.	A; IM; DM; L; V	A; M; L; S; EF	LoC; O; A; M; L; C; Cal; R
Reliability (Cronbach's α)	-	IM: 0.75; V: 0.76; L 0.33; A: 0.16; DM: 0.77; Total: 0.84	A 0.39, L: 0.4, M: 0.42, S: -0.14, EF: -0.37. Total: 0.60	0.71 person separation index 0.90 item separation index
Classification statistics	Sensitivity: 100%, 72%, 56%	Total Score: Sensitivity: 0.82 Specificity: 0.94 LR:13.7 PPV; NPV; OCC: 95.5; 78; 87.1	-	-

Construct validity	-	Partial evidence for convergent & divergent validity, except for 'A'. PCA: 2-factors	Partial evidence for convergent & divergent validity, except for EF.	Partial evidence for convergent validity, except for 'R'.
Criterion validity	-	Predictive: predicts FIM-cog & FAI scores. Concurrent: significant different scores left/right hemisphere stroke	Concurrent: Associated with FIM and MPAI-4	-

A= Attention, O= Orientation, M= Memory, VF= Verbal Fluency, L= Language, V= Visuospatial, IM= Immediate Memory, DM= Delayed Memory, S= Spatial, EF= Executive Functioning, LoC= Level of Consciousness, C= Construction, Cal= Calculation, R= Reasoning, LR= Likelihood Ratio, PPV= Positive Predictive Value, NPV= Negative Predictive Value, OCC= Overall Classification Correct, PCA= Principal Component Analysis, FIM-Cog= Functional Independence Measure-Cognitive, FAI = Frenchay Activity Index, MPAI-4= Mayo-Portland Adaptability Inventory-4.

ACE-R

Design & development:

The ACE was designed to screen for and distinguish between frontotemporal dementia and Alzheimer's disease (Mathuranath, Nestor, Berrios, Rakowicz, & Hodges, 2000). It was later revised (ACE-R) in order to improve ease of administration, sensitivity, ceiling effects, and to create parallel versions (Mioshi et al., 2006). A substantial body of research exists demonstrating the clinical utility and validity of the ACE-R for diagnosing dementia (for a review see Queally, Evans, & McMillan, 2011). The test also incorporates the Mini Mental State Examination (MMSE: Folstein, Folstein, & McHugh, 1975). It does not require additional equipment.

ABI studies:

One validation study has been conducted using the ACE-R in a working-age TBI population (Gaber, 2008). The sensitivity for both the ACE-R and MMSE scores were calculated at three different cut-offs drawn from the normative scores published in Mioshi et al., (2006). The ACE-R demonstrated better sensitivity at each cut-off compared to the MMSE. T-tests showed significant ($p<0.001$) differences between the MMSE, ACE-R total and ACE-R domain scores in this study and the Mioshi et al., (2006) normative data. One assumes that the ABI group was more impaired than the normative sample, though the paper does not state this. The authors conclude that the MMSE is not suitable for cognitive screening in ABI due to its insensitivity in detecting gross cognitive impairment but that the ACE-R has adequate sensitivity and can provide useful clinical information about the specific nature of difficulties.

Summary & critique

The authors acknowledge that the normative sample used for comparison on the current study had an older mean age (64.4 years) and more years of education (12.7). Whilst significant age effects have not been found in other studies (Mioshi et al., 2006; Larner, 2007), this may be due to restricted range and small sample sizes. Given that age effects are well established in other NTs (Lezak et al., 2004), obtaining normative date for a younger population seems necessary. The current study provides initial support for the practicality of using the ACE-R in a rehabilitation setting. However, it is limited in that it does not include a non-ABI control group to calculate specificity or predictive power.

RBANS

Design & development

The RBANS was also designed for the purpose of diagnosing and differentiating different subtypes of dementia in older adults and numerous studies have demonstrated its validity for this task (e.g. Randolph, 1997; 1998; Randolph, Tierney, Mohr, & Chase, 1998). Validation studies have also been conducted with other populations (e.g. Beatty et al., 2003; Aupperle, Beatty, Shelton, & Gontkovsky, 2002; Moser m& Schatz, 2002). The RBANS comprises 12 subtests which form five indices. It can be done at bedside and alternate forms are available.

ABI Studies

Reliability

One study has reported the internal reliability (Cronbach's α) of the RBANS in a TBI sample (McKay, Casey, Wertheimer, & Fichtenberg, 2007). As hypothesised, strong internal reliability was found for the Immediate/Delayed Memory, Visuospatial, and Total Scale Index scores. Weak internal reliability was found for Attention and Language Index scores as the subtests that make up these indexes are known to have differential sensitivities to the detection of brain damage (McKay et al., 2007).

Classification accuracy

The sensitivity of the RBANS to detect cognitive changes following a stroke was initially demonstrated by Duff, Beglinger, Jenks-Kettmann, and Bayless (2006) in their case study of a 22 year old woman with a complex psychiatric history who had suffered a right-hemisphere stroke. The RBANS and a battery of other NTs were administered pre and post-stroke and standardised change scores calculated. The largest decline was in the Visuospatial Index, in keeping with research on right hemisphere strokes (e.g. Jordan & Hillis, 2005). A similar pattern of changes was observed with the other NTs supporting the convergent validity of the RBANS. However, Reliable Change Index (RCI: Jacobson & Truax, 1991) scores are not reported so the significance of the change scores is unclear.

McKay, Wertheimer, Fichtenberg, and Casey (2008) compared performance on the RBANS between TBI patients and a matched clinical comparison group without ABI. They calculated the sensitivity (true positive rate), specificity (true negative rate), likelihood ratio (LR:

sensitivity/1-specificity), positive predictive value (PPV: proportion of positives that are true positives), negative predictive value (NPV: proportion of negatives that are true negatives) and overall correct classification (OCC: proportion correctly classified both positive and negative) at two cut-offs for RBANS Index and Total scores. Sensitivity and LRs ranged from modest to strong whilst specificity was high. Best performance was found for the Total score, with an OCC of 87.1%. The Attention Index was particularly sensitive, but had low specificity; the reverse pattern was true for the Visuospatial Index. An explanation of classification statistics would facilitate comprehension for readers unfamiliar with these.

Construct validity

The construct validity of the RBANS with ABI, stroke, and TBI patients is partially supported by Pachet (2007), Larson, Kirschner, Bode, Heinemann, and Goodman (2005) and McKay et al., (2007) respectively. Pachet (2007) found significant correlations between RBANS subtests and all matched NTs except RBANS Figure Copy/Recall and the Rey Complex Figure Test (RCFT: Meyers & Meyers, 1995). Only the Wechsler Memory Scale – III Digit Span correlated with RBANS Attention Index. Language subtests and Line Orientation were not included in this analysis, with no clear justification for this. Pachet (2007) concluded that the RBANS was not a strong measure of visual memory and that the Figure Copy/Recall and Coding subtests predominantly measure motor skills. McKay et al., (2007) similarly found moderate to strong correlations between RBANS subtests and NTs matched on basis of content and construct, with the exception of RBANS Figure Copy and the Benton Visual Retention Test (BVRT: Benton, 1974).

Larson et al., (2005) found that all RBANS Indices except Attention correlated significantly with NTs purported to measure the same construct, even whilst controlling for WAIS-R vocabulary scores (construed as a measure of general intelligence). The Attention Index showed a non-significant trend towards correlating with other tests of attention but was more strongly correlated with language tests, implying that this index may be a measure of general intelligence as opposed to attention. There was less evidence for divergent validity as the Attention, Visuospatial and Immediate Memory indices also correlated significantly with other language subtests, highlighting the RBAN's reliance on verbal abilities. Importantly, the authors controlled for multiple comparisons using Bonferroni corrections, reducing the possibility of making a Type I error. However, their sample size was smaller than originally reported as only 70/158 completed all tests, reducing statistical power.

Wilde (2006) also investigated RBANS construct validity through a principal components analysis which resulted in a two-factor solution accounting for 61% of the variance. This was termed a Language/Verbal Memory factor and a Visuospatial/Visual Memory factor. External validity of the factors was supported by significant correlations with NTs tapping similar constructs, as well as by the expected finding that left hemisphere stroke patients performed significantly worse on the Language factor and right hemisphere stroke patients performed significantly worse on the Visuospatial factor. The authors suggest that the five-factor structure of the RBANS may not be valid for right and left hemisphere stroke patients for whom language and visuospatial impairments may obscure deficits in memory and attention (Wilde, 2006). Wilde (2010) expanded on this finding by examining the different subtest and Index scores for left, right, bilateral, cortical and subcortical stroke patients using multivariate analysis of variance (MANOVA). As expected, left hemisphere patients were more impaired on Language; right hemisphere patients were more impaired on Visuospatial.

However, the results also suggested that impairments on Memory and Attention may be largely due to language deficits and the authors concluded that Attention Index has limited utility for stroke patients. The follow up study encompassed participants from the original study, therefore one might expect the same pattern of results to emerge. A validation study on an independent sample is needed to replicate these findings.

Predictive validity

Regarding predictive validity, Larson et al., (2005) found that the RBANS Total and Index scores (except Attention) at admission predicted cognitive outcome, as measured by the Functional Independence Measure (FIM: Hamilton, Granger, Sherwin, Zielzny & Tashman, 1987), at 12 month follow-up. The Visuospatial Index was also found to significantly correlate with activities of daily living, as measured by the Frenchay Activity Index (FAI: Holbrook & Skilbeck, 1983). Motor functioning, as measured by the motor scale of the FIM, and participation restrictions, as measured by the Craig Handicap and Assessment Reporting Technique (CHART: Whiteneck, Charliufe, Gerhart, Overholser, & Richardson, 1992), were not predicted by RBANS Total or Index scores. Attrition in the study was high (over 50%) resulting in a small sample size, therefore regression analyses could not be conducted limiting the conclusions that can be drawn.

Summary

The above studies provide preliminary evidence for the internal reliability, diagnostic validity and construct validity of the RBANS in ABI populations. However, as the RBANS was designed to screen for dementia, the individual subtests may not be the most appropriate for

ABI, particularly those comprising the Attention Index. It appears that the five-factor structure of the RBANS is also not valid, at least for stroke patients, with a two-factor verbal and visual solution being more appropriate. Many of the studies recommend supplementing the RBANS with additional measures of attention and executive function in ABI; however this introduces the issue of comparing tests with different normative samples.

NAB-SM

Design & development

In contrast to the tests reviewed so far, the NAB was developed to assess the cognitive abilities in adults aged 18 to 97 with the aim of detecting any central nervous system disorder. The full battery consists of 6 modules: a Screening Module and five domain-specific modules. The NAB-SM contains identical or similar items to the NAB and was designed to be able to predict performance on the NAB. The NAB combines the advantages of both a fixed and flexible battery due to its co-norming approach, which means that the NAB-SM can be administered on its own or together with other modules as indicated. T-scores can be calculated for subtest, index and total NAB-SM scores and it has a parallel form (White & Stern, 2003).

ABI studies

Reliability

Zgaljardic and Temple (2010) evaluated the internal consistency of the NAB-SM in an ABI sample. Cronbach's α was weak for each cognitive domain and satisfactory (0.60) for the

Total score. This suggests that clinicians should refer to subtest scores for interpretation, as Index scores may not be reliable indicators of their component subtests. One issue is that the study uses both versions of the NAB-SM with no separate analysis of each, so form equivalency cannot be assessed.

Construct validity

Zgaljardic and Temple (2010) computed correlations between NAB-SM Total and Index scores and NTs. Convergent and divergent validity was largely established, particularly for the Attention Index. However, the Executive Functions Index correlated with tests sharing a visuomotor component rather than tests of executive function. They also calculated the correlations between the NAB-SM subtests and matched and unmatched (in terms of content) NTs. Significant correlations were found with all matched tests except Delayed Shape Learning and Wechsler Memory Scale-3 (WMS-III) Visual Reproduction II. All unmatched NTs correlated with one or more NAB-SM subtest, with the exceptions of NAB-SM Digit Forward and NAB-SM Visual Discrimination. No corrections for multiple comparisons were made, increasing the possibility of making a Type I error. The authors concluded that the NAB-SM may be limited in screening for visual memory and executive function difficulties in an ABI population.

Ecological validity

Temple et al., (2009) and Zgaljardic, Yancy, Temple, Watford, and Miller (2011) investigated the ecological validity of the NAB-SM in a TBI sample by examining its relationship with the FIM and Mayo-Portland Adaptability Inventory-4 (MPAI-4: Malec &

Lezak, 2003) respectively. Hierarchical regression analysis controlling for age and gender demonstrated that the NAB-SM total score was significantly associated with the FIM total, motor and cognitive scores, accounting for 26%, 11% and 53% of the variance respectively (Temple et al., 2009). The NAB-SM Spatial Index was also significantly associated with the FIM total, cognitive and motor scores; whilst NAB-SM Language and Memory Index scores were significantly associated with the FIM cognitive score only. The NAB-SM Attention and Executive Function Indices did not correlate with the FIM.

Zgaljardic et al., (2011), using linear regression analysis, also found that the NAB-SM Total score was significantly associated with the MPAI-4 Total, Ability, and Participation scores, accounting for 9%, 9%, and 13% of the variance respectively. NAB-SM Spatial Index was also significantly correlated with the MPAI-4. However, in contrast to Temple et al., (2009), the Language and Memory Index scores did not correlate with the MPAI-4, whilst the Executive Function Index did. Divergent validity was also supported by the lack of significant associations between the NAB-SM and the MPAI-4 Adjustment score, a measure of affect.

Summary and critique

Whilst the NAB-SM Total score has adequate reliability, the Index scores do not which limits their clinical utility for cognitive profiling. The Executive Function Index in particular appears to have limited validity and may be measuring visuomotor processing skills. As executive impairments are common in TBI this is a major limitation of the screen. It appears that the NAB-SM is measuring constructs that map onto daily life functioning, which is

important in terms of informing rehabilitation, although the predictive validity of this is yet to be established through longitudinal studies.

Cognistat

Design & development

The Cognistat was designed to screen for cognitive impairment across a range of neurological or neuropsychiatric disorders (Kiernan et al., 1987). The test employs a screen and metric approach whereby for each domain a challenging question is first administered and if this is passed a full score is given for that domain, with no need for further testing. However, Oehlert et al., (1997) suggested this approach may underestimate impairment and recommended administering the entire test. Index scores are calculated and rated as average or mildly/moderately/severely impaired. A Total score can also be calculated but is not rated. Due to equipment requirements patients must be sat upright.

ABI Studies

Measurement properties:

Rasch rating scale analysis procedures (Rasch, 1980) were employed to evaluate 1) unidimensionality, 2) ability to characterise a useful range of cognitive function, and 3) the pattern of performance on cognitive domains of the Cognistat in a community TBI sample (Doninger et al., 2000). A series of calibrations resulted in three strata of performance, with the easiest item being Naming Objects and hardest being Delayed Recall. Rasch analysis suggested that the test was too easy for this sample and there were three misfitting items

which may be measuring alternative domains. A second analysis was run with a reduced item-set which created a less skewed test without increasing error. Analysis of the cognitive domains found that the Attention Index performed differently to the other domains, suggesting it may not be a good measure of general cognitive capacity.

Doninger et al., (2006) replicated and extended this study to compare findings between inpatient and community TBI populations. Results were comparable between the two samples, with the exception of less pronounced ceiling effects in the inpatient group. Due to the large and overlapping range of injury onset and severity, the two groups may not be as distinct as implied. The authors conclude that whilst the Cognistat may be useful as a crude measure of general cognitive dysfunction, or to reveal unexpected performance patterns, it is inappropriate to use for cognitive profiling in either community or inpatient TBI populations.

Construct validity

Nabors et al., (1997) found significant correlations between Cognistat subtests and associated NTs with the exception of the Reasoning subtests and Wisconsin Card Sorting Test (WCST: Heaton, 1981). However, they followed the screen and metric approach to administration which may underestimate impairment and inflate correlations. Wallace, Caroselli, Scheibel, and High (2000) reported the same pattern of findings, though with slightly different matched NTs. They also reported a significant correlation between the number of impaired performances on the Cognistat and NTs. Overall classification agreement between the Cognistat and NTs was fair (0.79 accuracy), with high sensitivity (0.92) but low specificity (0.22). However, at the subtest level classification agreement was interpreted as good for

Construction subtest only, all remaining subtests were in the poor range. The authors concluded that whilst the Cognistat may be useful in detecting the presence or absence of cognitive impairment, it should not be used for the purpose of describing an individual's cognitive profile.

Summary and critique

The above studies highlight concerns with the measurement properties of the Cognistat when used with ABI populations. The test demonstrates ceiling effects, appears to have many redundant or poorly constructed items, and the validity of the screen and metric approach has been questioned. Whilst the measure may be able to differentiate three levels of cognitive function at a global level, it seems inappropriate for use in cognitive profiling.

Discussion

This review aimed to identify and evaluate the psychometric properties of brief but comprehensive NTs. Four measures were identified and reviewed. Whilst all measures have some preliminary evidence for their use in cognitive assessment of ABI, research is lacking and each has limitations.

The ACE-R and RBANS were both designed specifically as a tool for screening for dementia. This is important in terms of the content validity of the measures, as the theoretical model of cognitive impairment in dementia differs from that for ABI. For example, in Alzheimer's disease the temporal lobes are commonly affected, which is associated with

deficits in category fluency (Monsch et al., 1992); whilst in TBI frontal lobe damage is more common, resulting in deficits in letter fluency (Cummings, 1995). Therefore, the rationale behind test item selection will differ and items included in the RBANS and ACE-R may not be the most suitable for the purpose of cognitive assessment in ABI.

Of the four measures reviewed, the NAB-SM provides the largest and most comprehensive normative data. Sample sizes for the normative data for both the ACE-R and Cognistat are relatively small, increasing the possibility of error and limiting generalizability. Scores are not scaled according to age, gender, or education level and the ACE-R normative sample are aged over 50 years, making comparisons in younger individuals problematic. As normative comparison is a fundamental principal in neuropsychological assessment (Harvey, 2012), having representative norms that have been stratified according to variables known to influence test performance is essential.

All of the measures include an attention, memory, language and visuospatial/constructional component. The Cognistat also includes a Calculation and Reasoning Index, though the construct validity of the latter was not supported by the evidence from Nabors et al., (1997). Only the NAB-SM includes an Executive Function Index. Executive impairments are one of the most common sequelae of ABI (Lesniak, Bak, Czepiel, Seniow & Czlonkowska, 2008) and are associated with increased dependence (Lesniak et al., 2008), poorer return to work (Ownsworth & Shum, 2008) and decreased social participation (McDowd, Filion, Pohl, Richards, & Stiers, 2003). Therefore it seems essential for any NT to include items purported to measure this construct. Whilst it is possible to supplement assessment with tests of

executive function such as the WCST, this introduces the problem of normative comparison and increases the chance of making invalid interpretations (Russell et al., 2005).

Evidence for strong psychometric properties in an ABI sample is not well established for any of the measures reviewed. No studies have examined the reliability of the ACE-R. Cronbach's α statistics were acceptable for the RBANS Immediate/Delayed Memory Index, Visuospatial Index, and Total Scale Index only (McKay et al., 2007). No α's reached acceptable levels in the NAB-SM (Zgaljardic & Temple, 2010). This suggests that the Attention and Language RBANS indices and the NAB-SM Indices may not be valid indicators of the underlying subtest performance. Similarly, Doninger et al., (2000; 2006) found that the Cognistat could only reliably differentiate three different levels of performance at the global level, not cognitive domain level.

The fact that information about classification accuracy is currently unavailable for ACE-R, NAB-SM and Cognistat significantly limits the clinical utility of these measures with ABI populations. Classification accuracy statistics were acceptable for the RBANS Total score (McKay et al., 2008) but not for the individual Index scores. This is not an issue for diagnosis of brain dysfunction (as individual indices would not be used for this purpose). However, it would be useful to know the classification agreement between the Index scores on these measures and corresponding scores on gold standard NTs. This information is only reported for the Cognistat (Wallace et al., 2000), with poor agreement found between most items.

In general, more evidence is needed for both construct and criterion related validity for all the included measures. Whilst significant correlations have been found between most subtests and indices and other matched NTs, providing evidence for convergent validity, this is weaker for the RBANS Attention Index, NAB-SM Executive Function Index, and Cognistat Reasoning Index. Evidence for divergent validity is more variable. This may indicate a shared variable (e.g. general intelligence) across many items. Predictive validity has only been evaluated with the RBANS (Larson et al., 2005). Though the NAB-SM has been shown to relate to functional abilities concurrently (Temple et al., 2009; Zgaljardic et al., 2011), longitudinal studies are needed to evaluate the predictive validity of this measure.

The RBANS or the NAB-SM seem most suitable for assessing cognitive impairment in ABI. RBANS has the advantage of better reliability and a more extensive evidence base including classification statistics and predictive validity. However, it was originally designed as a dementia screen; its five-factor structure is not supported; and it does not measure executive function. The NAB-SM has the advantage of being designed as a generic screen and including an Executive Function Index. However, the construct validity of this Index has been questioned; reliability is poor; and no information is available on classification accuracy.

Future research

One avenue for future research is to add to the existing psychometric evidence base for the RBANS and NAB-SM. Evidence of other forms of reliability is needed. Information about the classification accuracy is required for the NAB-SM. Construct validity could be further

assessed by investigating the relationship between the NAB-SM and the NAB. Validation and comparison against a wider variety of other NTs, functional abilities or measures, and outcome measures is also indicated, as well as comparison of the RBANS and NAB-SM themselves. Longitudinal studies to assess change over time, practice effects and relationship with long-term outcomes would also further the evidence base. All of the above could be done with a variety of ABI populations to investigate the impact of, for example, cause and severity of injury.

A second avenue for future research would be to design and develop a new tool for assessing cognitive abilities in ABI. The advantage of this avenue is that the tool could be purpose-designed for ABI populations, with test items selected based on theoretical models of ABI and the existing literature on known cognitive sequelae. The test would need to be designed with the aims of addressing the limitations of the existing measures, for example reducing floor and ceiling effects, providing adequate normative data, and demonstrating acceptable reliability and validity.

Conclusions

This review has identified and critically evaluated the psychometric properties of four cognitive assessments used in ABI. It has concluded that the RBANS and NAB-SM show the most promise for this purpose, but both have significant limitations. Future research should aim to increase the evidence base for either of these measures with ABI. Alternatively, a new ABI-specific tool could be developed that aims to address the limitations of the existing tools.

References

Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Aupperle, R. L., Beatty, W. W., DeNap Shelton, F., & Gontkovsky, S. T. (2002). Three screening batteries to detect cognitive impairment in multiple sclerosis. Multiple Sclerosis, 8(5), 382-389.

Beatty, W. W., Ryder, K. A., Gontkovsky, S. T., Scott, J. G., McSwan, K. L., & Bharucha, K. J. (2003). Analyzing the subcortical dementia syndrome of Parkinson's disease using the RBANS. Archives of Clinical Neuropsychology, 18, 509–520.

Benton, A. L. (1974). Visual retention test. Psychological Corporation.

Chelune, G. J. (2010). Evidence-based research and practice in clinical neuropsychology. The Clinical Neuropsychologist, 24(3), 454-467.

Cullen, B., O'Neill, B., Evans, J. J., Coen, R. F., & Lawlor, B. A. (2007). A review of screening tests for cognitive impairment. Journal of Neurology, Neurosurgery & Psychiatry, 78(8), 790-799.

Cummings, J. L. (1995). Anatomic and Behavioral Aspects of Frontal-Subcortical Circuitsa. Annals of the New York Academy of Sciences, 769(1), 1-14.

Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. Neuropsychology, 9(1), 80.

Dikmen, S., McLean Jr, A., Temkin, N. R., & Wyler, A. R. (1986). Neuropsychologic outcome at one-month postinjury. Archives of Physical Medicine Rehabilitation, 67(8), 507-513.

Doninger, N. A., Ehde, D. M., Bode, R. K., Knight, K., Bombardier, C. H., & Heinemann, A. W. (2006). Measurement properties of the neurobehavioral cognitive status examination (Cognistat) in traumatic brain injury rehabilitation. Rehabilitation Psychology, 51(4), 281.

Doninger, N. A., Bode, R. K., Heinemann, A. W., & Ambrose, C. (2000). Rating scale analysis of the Neurobehavioral Cognitive Status Examination. Journal of Head Trauma Rehabilitation, 15, 683–695.

Duff, K., Beglinger, L. J., Jenks Kettmann, J. D., & Bayless, J. D. (2006). GRAND ROUNDS: Pre-and Post-Right Middle Cerebral Artery Stroke in a Young Adult: A Case Study Examining the Sensitivity of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). Applied neuropsychology, 13(3), 194-200.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research, 12(3), 189-198.

Fritz, J. M., & Wainner, R. S. (2001). Examining diagnostic tests: an evidence-based perspective. Physical Therapy, 81(9), 1546-1564.

Gaber, T. A. Z. (2008). Evaluation of the Addenbrooke's Cognitive Examination's validity in a brain injury rehabilitation setting. Brain Injury, 22(7-8), 589-593.

Greenhalgh, T. (1997). How to read a paper. Papers that report diagnostic or screening tests. BMJ: British Medical Journal, 315(7107), 540.

Hamilton, B. B., Granger, C. V., Sherwin, F. S., Zielzny, M., & Tashman, J. S. (1987). A uniform national data system for medical rehabilitation. In Fuhrer, M. J. (Ed.), Rehabilitation outcomes: Analysis and measurement. Baltimore: Brookes, p. 137–147.Harvey, 2012

Harvey, P. D. (2012). Clinical applications of neuropsychological assessment.Dialogues in clinical neuroscience, 14(1), 91.

Headway: the brain injury association. About brain injury. Retrieved from https://www.headway.org.uk/About-Brain-Injury.aspx

Heaton, R. K. (1981). A manual for the Wisconsin card sorting test. Western Psycological Services.

Heilman, K. M., & Valenstein, E. (2003). Clinical neuropsychology. European Journal of Neurology, 10(5), 606-606.

Herrmann, L. L., Goodwin, G. M., & Ebmeier, K. P. (2007). The cognitive neuropsychology of depression in the elderly. Psychological medicine, 37(12), 1693-1702.

Hinkin, C. H., Castellon, S. A., Durvasula, R. S., Hardy, D. J., Lam, M. N., Mason, K. I., ... & Stefaniak, M. (2002). Medication adherence among HIV+ adults Effects of cognitive dysfunction and regimen complexity. Neurology, 59(12), 1944-1950.

Holbrook, M., & Skilbeck, C. E. (1983). An activities index for use with stroke patients. Age and ageing, 12(2), 166-170.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of consulting and clinical psychology, 59(1), 12.

Jaeschke, R., Guyatt, G., Sackett, D. L., & Evidence-Based Medicine Working Group. (1994). Users' guides to the medical literature: III. How to use an article about a diagnostic test: A. Are the results of the study valid?. JAMA-Journal of the American Medical Association-US Edition, 271(5), 389-391.

Jin, Y. P., Di Legge, S., Ostbye, T., Feightner, J. W., & Hachinski, V. (2006). The reciprocal risks of stroke and cognitive impairment in an elderly population. Alzheimer's & dementia, 2(3), 171-178.

Jordan, L. C., & Hillis, A. E. (2005). Aphasia and right hemisphere syndromes in stroke. Current neurology and neuroscience reports, 5(6), 458-464.

Kiernan, R. J., Mueller, J., Langston, J. W., & Van Dyke, C. R. A. I. G. (1987). The Neurobehavioral Cognitive Status Examination: A brief but differentiated approach to cognitive assessment. Annals of internal medicine, 107(4), 481-485.

Kneipp, S., & Rubin, A. (2007). Community re-entry issues and long-term care. Brain injury medicine, 1085-1104.

Larner, A. J. (2007). Addenbrooke's Cognitive Examination-Revised (ACE-R) in day-to-day clinical practice. Age and ageing, 36(6), 685-686.

Larson, E. B., Kirschner, K., Bode, R., Heinemann, A., & Goodman, R. (2005). Construct and predictive validity of the Repeatable Battery for the Assessment of Neuropsychological Status in the evaluation of stroke patients. Journal of Clinical and Experimental Neuropsychology, 27(1), 16-32.

Leśniak, M., Bak, T., Czepiel, W., Seniów, J., & Członkowska, A. (2008). Frequency and prognostic value of cognitive disorders in stroke patients. Dementia and geriatric cognitive disorders, 26(4), 356-363.

Lezak, M. D. (Ed.). (2004). Neuropsychological asessment 4 Ed. Oxford university press.

Malec, J. F., & Lezak, M. D. (2003). Manual for the mayo-portland adaptability inventory (MPAI-4). Portland: Oregon Health and Sciences University.

Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., & Hodges, J. R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. Neurology Minneapolis, 55(11), 1613-1620.

McCue, M., Rogers, J. C., & Goldstein, G. (1990). Relationships between neuropsychological and functional assessment in elderly neuropsychiatric patients. Rehabilitation Psychology, 35(2), 91.

McDowd, J. M., Filion, D. L., Pohl, P. S., Richards, L. G., & Stiers, W. (2003). Attentional abilities and functional outcomes following stroke. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58(1), P45-P53.

McKay, C., Casey, J. E., Wertheimer, J., & Fichtenberg, N. L. (2007). Reliability and validity of the RBANS in a traumatic brain injured sample. Archives of Clinical Neuropsychology, 22(1), 91-98.

McKay, C., Wertheimer, J. C., Fichtenberg, N. L., & Casey, J. E. (2008). The repeatable battery for the assessment of neuropsychological status (RBANS): clinical utility in a traumatic brain injury sample. The Clinical Neuropsychologist, 22(2), 228-241.

McMillan, T. M., & Ledder, H. (2001). A survey of services provided by community neurorehabilitation teams in South East England. Clinical Rehabilitation, 15(6), 582-588.

Meyers, J. E., & Meyers, K. R. (1995). Rey complex figure test and recognition trail. Psychological Assessment Resources Incorporated.

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. International journal of geriatric psychiatry, 21(11), 1078-1085.

Monsch, A. U., Bondi, M. W., Butters, N., Salmon, D. P., Katzman, R., & Thal, L. J. (1992). Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. Archives of Neurology, 49(12), 1253.

Moser, R. S., & Schatz, P. (2002). Enduring effects of concussion in youth athletes. Archives of Clinical Neuropsychology, 17(1), 91-100.

Nabors, N. A., Millis, S. R., & Rosenthal, M. (1997). Use of the Neurobehavioral Cognitive Status Examination (Cognistat) in traumatic brain injury. The Journal of Head Trauma Rehabilitation, 12(3), 79-84.

Oehlert, M. E., Hass, S. D., Freeman, M. R., Williams, M. D., Ryan, J. J., & Sumerall, S. W. (1997). The neurobehavioral cognitive status examination: Accuracy of the “screen-metric” approach in a clinical sample. Journal of clinical psychology, 53(7), 733-737.

Ownsworth, T., & Shum, D. (2008). Relationship between executive functions and productivity outcomes following stroke. Disability & Rehabilitation, 30(7), 531-540.

Pachet, A. K. (2007). Construct validity of the Repeatable Battery of Neuropsychological Status (RBANS) with acquired brain injury patients. The Clinical Neuropsychologist, 21(2), 286-293.

Ponsford, J., Draper, K., Schonberger, M., Baddeley, A., Emslie, H., Nimmo-Smith, I., ... & Franulic, A. (2008). Functional outcome 10 years after traumatic brain injury: Its relationship with demographic, injury severity, and cognitive and emotional status. Journal of the International Neuropsychological Society, 14(2), 233.

Poulin, V., Korner-Bitensky, N., & Dawson, D. R. (2013). Stroke-specific executive function assessment: A literature review of performance-based tools. Australian occupational therapy journal, 60(1), 3-19.

Queally, V. R., Evans, J. J., & McMillan, T. M. (2011). A meta-analysis of studies comparing the effectiveness of three cognitive screening tests in the detection of dementia populations. International journal of geriatric psychiatry, 26(5), 548-549.

Randolph, C. (1997). Differentiating vascular dementia from alzheimer's disease: The role of neuropsychological testing. Clinical Geriatrics, 5, 77–86.

Randolph, C. (1998). Repeatable battery for the assessment of neuropsychological status: Manual. San Antonio, TX: The Psychological Corporation

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. Journal of clinical and experimental neuropsychology, 20(3), 310-319.

Rasch, W. (1980). Die psychologisch-psychiatrische Beurteilung von Affektdelikten. NJW, 24, 1309-1315.

Russell, E. W., Russell, S. L., & Hill, B. D. (2005). The fundamental psychometric status of neuropsychological batteries. Archives of Clinical Neuropsychology, 20(6), 785-794.

Sarti, C., Rastenyte, D., Cepaitis, Z., & Tuomilehto, J. (2000). International trends in mortality from stroke, 1968 to 1994. Stroke, 31(7), 1588-1601.

Sawrie, S. M., Martin, R. C., Gilliam, F. G., Roth, D. L., Faught, E., & Kuzniecky, R. (1998). Contribution of neuropsychological data to the prediction of temporal lobe epilepsy surgery outcome. Epilepsia, 39(3), 319-325.

Schanke, A. K., & Sundet, K. (2000). Comprehensive Driving Assessment: Neuropsychological Testing and On-road Evaluation of Brain Injured Patients. Scandinavian journal of psychology, 41(2), 113-121.

Schoenberg, M. R. (2011). The Little Black Book of Neuropsychology: A Syndrome-based Approach. M. R. Schoenberg, & J. G. Scott (Eds.). Springer.

Schanke, A. K., & Sundet, K. (2000). Comprehensive Driving Assessment: Neuropsychological Testing and On-road Evaluation of Brain Injured Patients. Scandinavian journal of psychology, 41(2), 113-121.

Shames, J., Treger, I., Ring, H., & Giaquinto, S. (2007). Return to work following traumatic brain injury: trends and challenges. Disability & Rehabilitation, 29(17), 1387-1395.

Stern, R. A., & White, T. (2003). Neuropsychological Assessment Battery. Lutz, FL: Psychological Assessment Resources.

Temple, R. O., Zgaljardic, D. J., Abreu, B. C., Seale, G. S., Ostir, G. V., & Ottenbacher, K. J. (2009). Ecological validity of the neuropsychological assessment battery screening module in post-acute brain injury rehabilitation. Brain injury, 23(1), 45-50.

Wallace, J. J., Caroselli, J. S., Scheibel, R. S., & High Jr, W. M. (2000). Predictive validity of the Neurobehavioural Cognitive Status Examination (NCSE) in a post-acute rehabilitation setting. Brain injury, 14(1), 63-69.

Wallesch, C. W., Curio, N., Galazky, I., Jost, S., & Synowitz, H. (2001). The neuropsychology of blunt head injury in the early postacute stage: effects of focal lesions and diffuse axonal injury. Journal of Neurotrauma, 18(1), 11-20.

White, T., & Stern, R. A. (2003). NAB, Neuropsychological Assessment Battery: Psychometric and Technical Manual. Psychological Assessment Resources.

Whiteneck, G.G., Charliufe, S.W., Gerhart, K.A., Overholser, J.D., & Richardson, G.N. (1992). Guide for use of the CHART: Craig Handicap Assessment and Reporting Technique. Authors: Englewood, Colorado.

World Health Organisation (WHO). (2006). Neurological disorders: Public health challenges. Retrieved from: http://www.who.int/mental_health/neurology/neurological_disorders_report_web.pdf

Wilde, M. C. (2006). The validity of the repeatable battery of neuropsychological status in acute stroke. The Clinical Neuropsychologist, 20(4), 702-715.

Wilde, M. C. (2010). Lesion location and repeatable battery for the assessment of neuropsychological status performance in acute ischemic stroke. The Clinical Neuropsychologist, 24(1), 57-69.

Wood, R., Alderman, N., & Williams, C. (2008). Assessment of neurobehavioural disability: A review of existing measures and recommendations for a comprehensive assessment tool. Brain Injury, 22(12), 905-918.

Zgaljardic, D. J., & Temple, R. O. (2010). Reliability and validity of the Neuropsychological Assessment Battery-Screening Module (NAB-SM) in a sample of patients with moderate-to-severe acquired brain injury. Applied neuropsychology, 17(1), 27-36.

Zgaljardic, D. J., Yancy, S., Temple, R. O., Watford, M. F., & Miller, R. (2011). Ecological validity of the screening module and the Daily Living tests of the Neuropsychological

Assessment Battery using the Mayo-Portland Adaptability Inventory-4 in postacute brain injury rehabilitation. Rehabilitation psychology, 56(4), 359.

Section B:

The reliability, validity and factor structure of the Short Parallel Assessments of Neuropsychological Status (SPANS)

Jennifer Attwood

July 2013

Word count:

7493

A thesis submitted in partial fulfilment of the requirements of

Canterbury Christ Church University for the degree of

Doctor of Clinical Psychology

Abstract

Cognitive complaints are common following an acquired brain injury and require careful assessment in order to guide treatment and care. There is a need for brief, comprehensive and psychometrically valid tests of cognitive function that can be used in neuro-rehabilitation services by a range of health professionals. The Short Parallel Assessment of Cognitive Status (SPANS) was purpose-designed to meet this need. The current study assessed the reliability, discriminative validity and factor structure of the SPANS. Participants were 61 people with an acquired brain injury, 35 people with a long-term neurological condition, and 122 healthy controls. Cronbach's alphas were adequate to excellent for the clinical groups though poor for the healthy controls due to limited variance in the scores. Receiver operating characteristic curves showed that SPANS indices were significantly able to discriminate between people with a neurological condition and healthy controls as well as between left and right hemisphere damage. Exploratory factor analysis suggested the retention of 25 subtests representing three factors that largely followed the purported structure of the test: Memory and Learning, Language, and Visual-motor Performance. Limitations of the study, clinical/theoretical implications and research directions are considered. It is concluded that the SPANS is a reliable and valid tool for the assessment of cognitive function in people with an acquired brain injury, though further validation studies are required.

Keywords: Neuropsychological assessment, screening tool, cognitive impairment, psychometrics, acquired brain injury

Introduction

Acquired brain injury (ABI) is the term used for any injury to the brain that has occurred since birth and is non-degenerative (Headway: the brain injury association). People present to services with a complex array of difficulties following ABI, including physical disability, cognitive impairment, behavioural problems and issues with social and daily-life functioning (Turner-Stokes, Nair, Sedki, Disler, & Wade, 2011). Cognitive impairment is common and may include difficulties with memory, attention, processing speed, language, visuo-spatial or perceptual abilities, and executive function (Williamson, Scott, & Adams, 1996; Hanks, Ricker, & Millis, 2004; King & Tyerman, 2003; Lesniak, Bak, Czepiel, Seniow, & Czlonkowska, 2008). General theoretical frameworks in neuropsychology distinguish between cognitive functions that have a distributed neural basis, such as attention, memory, and executive functions, and cognitive functions that are more localised neuroanatomically, such as language (Hodges, 1999). Both diffuse and localised damage can occur following ABI resulting in complex symptomatic presentations (Kreutzer, Gordon, Rosenthal, & Marwitz, 1993).

Studies have shown that early cognitive status is associated with the ability to benefit from further rehabilitation, need for post-discharge support, and overall outcome (Dikmen, McLean, Temkin, & Wyler, 1986). The results of neuropsychological assessments can inform treatment planning and improve outcomes for patients and medico-legal clients (Chelune, 2010). Therefore, an important role for clinical psychologists working in neurorehabilitation services is the objective assessment of cognitive dysfunction using standardised neuropsychological measures (Harvey, 2012). Cognitive assessment needs not only to take account of the broad range of cognitive domains highlighted above, but also their constituent

parts. For example, Sohlberg and Mateer (1989) described five aspects of attention in order of increasing difficulty: focused, sustained, selective, alternating, and divided. Information processing theories of memory generally describe three key processes: encoding, consolidation and retrieval, which are further subdivided into different sensory modalities, such as verbal and visual (e.g. Schoenberg, 2011) Comprehensive assessment is therefore essential.

The most common approach to assessing cognition in clinical practice is using a flexible battery (i.e. a collection of individually standardised tests) to measure a range of different cognitive constructs (Kosaka, 2006; Rabin, Barr, & Burton, 2005; Sweet, Nelson, & Moberg, 2006). However, this approach has been criticised (Russell, Russell, & Hill, 2005) for not adhering to one of the fundamental principles of neuropsychological assessment, namely the reliability of measurement and the application of psychometric theory. Russell et al. (2005) argued that clinicians should therefore employ fixed or standardised batteries that have been co-normed and validated as a whole (for alternative views see Bigler, 2007 and Larrabee, 2008). Resources in public health services are limited. Patients are often unable to tolerate much testing due to factors such as fatigue and confusion. In the early stages of recovery (e.g. emerging from post-traumatic amnesia), health and cognitive status changes rapidly, rendering test results quickly obsolete. Together, this means that a test battery, as well as being psychometrically reliable and valid, also needs to be brief, repeatable and able to be delivered by a range of health professionals (Nabors, Millis, & Rosenthal, 1997). An accessible, comprehensive, co-normed and validated measure is required.

Review of existing brief comprehensive measures

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) was originally designed and validated as a screen for dementia (Randolph, 1997) but has since been used for a number of other clinical groups, including ABI. Internal reliability has been shown to be acceptable with a traumatic brain injury (TBI) sample for all indices, bar Language and Attention (McKay, Casey, Wertheimer, & Fichtenberg, 2007). Its classification accuracy (i.e. ability to diagnose cognitively impaired versus healthy controls) was adequate when using the Total Scale Index (McKay, Wertheimer, Fichtenberg, & Casey, 2008) but the individual indices were less discriminatory. Partial support for convergent and divergent validity with existing similar measures was found in a number of studies (Pachet, 2006; Larson, Kirschner, Bode, Heinemann, & Goodman, 2005; McKay et al., 2007), and the RBANS was shown to predict cognitive outcome and activities of daily living at 12 month follow up (Larson et al., 2005). However, as the test items were chosen with the aim of detecting dementia, not all items were deemed sensitive to the type of brain dysfunction commonly found following ABI (McKay et al., 2007) and important cognitive domains, such as executive functioning, are not assessed. The Attention and Visual Memory components of the RBANS have also been criticised as not measuring what they purport to (Pachet, 2006) and Wilde (2006) demonstrated with principal component analysis that a two-factor solution (verbal reasoning/memory and visual reasoning/memory) was more interpretable than the proposed five-factor structure of the test.

The Neuropsychological Assessment Battery Screening Module (NAB-SM) has the advantage of being designed to screen for a range of cognitive abilities related to any central nervous system condition (Stern & White, 2003). However, internal reliability has been

reported as inadequate for both domain and total index scores, suggesting that the test items may not be measuring unitary constructs (Zgaljardic & Temple, 2010). Evidence for convergent and divergent validity has been reported (Zgaljardic & Temple, 2010); although the Executive Function Index appeared to be more strongly associated with tests of visuomotor processing than traditional tests of executive function. There is partial evidence for ecological validity in terms of veridicality, or correlation with measures of real-world functioning (Chaytor & Schmitter-Edgecombe, 2003), as the NAB-SM correlates with other functional measures (Temple et al., 2009; Zgaljardic, Yancy, Temple, Watford, and Miller, 2011), with the exception of the Attention and Executive Function Indices. A significant limitation of the NAB-SM is the lack of information about its classification accuracy either in relation to brain dysfunction compared to healthy or other clinical groups, or in relation to results of the full NAB.

The Short Parallel Assessments of Neuropsychological Status (SPANS)

The SPANS (Burgess, 2013) is a newly developed brief but comprehensive neuropsychological battery which was designed with the aims of 1) addressing the design and psychometric limitations of existing test batteries used with working-age ABI populations, 2) creating a repeatable test that could provide information relevant to typical referral questions in neurorehabilitation services and 3) being practical for use in both in-patient and community settings.

Relevant cognitive functions to assess were identified through a thorough review of the empirical and theoretical literature on common cognitive sequelae following ABI. For

example, spatial and object perception have distinct neuroanatomical pathways, so screening for both was deemed necessary (Warrington & James, 1991). Language tends to be represented predominantly in the left hemisphere, whilst visuo-spatial abilities are predominantly lateralised to the right hemisphere (Kolb & Whishaw, 2003), therefore including both visual and verbal tests can inform on localisation of damage. Specific tests were chosen based on their evidenced clinical utility (e.g. discriminative validity), predictive power in regard to post-discharge outcomes, and ecological validity in terms of verisimilitude (face validity: Chaytor & Schmitter-Edgecombe, 2003). For example, a task such as learning a grocery list has obvious real-world implications, thereby increasing face validity and cooperation from patients. In terms of discriminative validity, research has shown that the most sensitive measures of ABI compared to normal functioning are more 'ability-focused' tests (Larrabee, 2008), such as verbal supraspan learning tests, memory tests such as Logical Memory or Visual Reproduction subtests of the Wechsler Memory Scales, or sequencing/processing speed tests such as Trail Making Test B (Powell, Cripe, & Dodrill, 1991; Larrabee, 2008; Dikmen, Machamer, Winn, & Temkin, 1995). In comparison, tests of more 'crystallised intelligence' or semantic knowledge, such as the Information subtest of the Wechsler Intelligence Scales, tend to be poorer discriminators (Capruso & Levin, 1992). Hence subtests were chosen to maximise the sensitivity of the SPANS.

The SPANS was developed in routine clinical practice which gives it the advantage of addressing both the practical needs (e.g. administrable at bedside) and the typical referral questions (e.g. determine cognitive strength and weaknesses, assess learning potential) received in neurorehabilitation services (for full details see Burgess, 2013). Initial evidence for the reliability and validity of the SPANS is reported in the test manual (Burgess, 2013). The current study aimed to extend these findings to further evaluate the SPANS.

Aims and Hypotheses

1) A design aim in the development of the SPANS was to create reliable and meaningful index scales (Burgess, 2013). Therefore, it was hypothesised that the SPANS index scales would demonstrate high internal consistency for the clinical participant groups, both separately and combined, and for each parallel form separately and combined. As the SPANS was designed to detect cognitive impairment as opposed to variability in cognitive skills within a healthy population, it was further hypothesised that internal consistencies would be low within the healthy control group (as a result of limited variance within the scores).

2) There is evidence that the SPANS was able to discriminate between individuals with an ABI or other neurological condition and healthy controls (Burgess, 2013). However, the current study aimed to explore this in more depth by establishing optimum cut-off scores and evaluating the classification accuracy of each index score. It was hypothesised that the SPANS index scores would demonstrate adequate sensitivity and specificity in detecting ABI. It was further hypothesised that the indices containing more 'ability-focused' tests of memory, learning or processing speed would demonstrate better discriminative power compared to indices containing more semantic-based knowledge, such as the Language Index (LAI) (Larrabee, 2008; Carpuso & Levin, 1992). . Finally, it was hypothesised that the LAI and Visual Performance Index (VPI) would discriminate between left and right hemisphere damage respectively.

3) The final aim of the study was to assess the construct validity of the SPANS through exploration of the underlying factor structure and evaluation of the internal consistency of

any emerging factors. Although the SPANS includes seven index scores, two of these contain subtests that overlap with other indices, therefore it was hypothesised that five factors would emerge.

Methodology

Participants

Participants were a subset[2] of those recruited by Burgess (2013). Clinical participants were 96 non-consecutive referrals to three neuropsychology services in the UK over a four-year period (2007-2011). The majority of these (n = 77) were given SPANS form A (as this was designed to be administered first); however a randomly selected subset (n = 19) received form B as part of the parallel test evaluation process. Participants with any type of ABI were included if they were able to complete the SPANS and cognitive assessment was warranted and relevant to the referral question. Patients with excessive fatigue, language, visual, or motor impairments were excluded. Demographic and injury related information was obtained by the multi-disciplinary team as part of usual clinical practice. The clinical sample was composed of two groups: 1) 61 patients with a diagnosis of an ABI within the last year (ABI) and 2) 35 patients with a long term neurological condition, including ABIs more than one year old and other non-ABI diagnoses such as multiple sclerosis (LTNC: see Table 1 for more details). Demographic information for the clinical and control groups can be seen in Table 2.

[2] Time two repeated measures data were not included hence the smaller sample size than Burgess (2013).

Table 1: Injury and severity information by clinical group and SPANS form

	ABI(n=61)		LTNC (n=35)	
Form	A (n=49)	B (n=12)	A (n=28)	B (n=7)
TBI	24	6	9	4
Anoxic/Hypoxic	5	4	0	2
Subarachnoid/subdural Haemorrhage	10	2	1	1
Stroke/Other Haemorrhage	6	0	3	0
Other ABI*	4	0	6	0
Epilepsy	0	0	2	0
Other Non-ABI**	0	0	6	0
LH/RH/BI/UN***	16/15/11/7	5/1/2/4	5/6/2/15	0/2/2/3
Lowest GCS****: Mean (SD)	7.10 (3.48)	8.00 (3.83)	9.18 (4.69)	9.43 (4.58)
Weeks since injury (ABI)/initial diagnosis (LTNC): Mean (SD) range	10.17 (8.56) 3-42	15.63 (15.94) 2-50	254.00 (145.89) 59-525	319.00 (182.09) 76-550

* 'Other ABI' = hydrocephalus, viral encephalitis, and profound hypoglycaemia.

** 'Other Non-ABI' = multiple sclerosis, brain tumour, developmental disorder, and unexplained conditions characterised by neuropsychological deficit.

*** LH = left hemisphere; RH = right hemisphere; BI = bilateral; UN = undifferentiated

****GSC = Glasgow Coma Scale (Teasdale & Jennett, 1974) = a measure of severity scored in the first days/weeks following ABI. A score of 13-15 represents a minimal-to-mildly severe ABI; 9-12 represents a moderately severe ABI; and 3-8 represents a severe ABI. This information was not collected for non-ABI patients.

Table 2: Demographic information by participant group and SPANS form

	ABI(n=61)		LTNC (n=35)		Controls (n=122)	
Form	A (n=49)	B (n=12)	A (n=28)	B (n=7)	A (n=61)	B (n=61)
Sex M/F	38/11	11/1	20/28	5/2	37/24	26/35
Age (yrs) mean	44.92	45.25	38.54	49.57	51.59	42.46
(SD)	(15.47)	(14.33)	(12.20)	(11.82)	(14.99)	(18.37)
Education						
Secondary	17	6	8	3	26	35
College/vocational	22	5	15	3	19	20
University	10	1	5	1	16	6
WTAR *	n=6	n=0	n=17	n=2		
Mean (SD)	103		98.29	112.5	103.43	101.70
	(14.17)		(13.51)	(4.95)	(5.18)	(5.13)

* WTAR = Wechsler Test of Adult Reading (Holdnack, 2001)

The normative sample consisted of 122 healthy controls recruited between 2009 and 2013, 50% of whom were randomly allocated to receive SPANS form B. The sample was composed of patient family members, NHS employees, and other contacts of the research team. Participants were included if they had a Wechsler Test of Adult Reading (WTAR: Holdnack, 2001) estimated Average IQ (score between 90 and 110). Participants were excluded if they had a history of ABI or other neurological condition. All spoke English as a first language and had been educated in an English-speaking, Western culture.

Measures

SPANS

As described, the SPANS is a brief but comprehensive neuropsychological test battery designed to assess cognitive abilities following ABI. It is composed of 33 subtests that make up seven indices: Orientation (ORI), Attention/Concentration (ACI), Memory/Learning (MLI), Language (LAI), Visual-Motor Performance (VPI), Processing Speed (PSI), and Conceptual Flexibility (CFI) (see Appendix C for list of subtests/indices). There are two parallel versions of the test, designed to be equivalent and to minimise practice effects. The test manual (Burgess, 2013) reported at least adequate internal consistencies (Cronbach's $\alpha \geq .7$; Kline, 1999) for the parallel versions and clinical/control groups combined. Despite high correlations between the index scores, suggesting they measure a shared ability, there remained unshared variance of at least 38.4%, implying that the tests also tap distinct abilities. The manual also reported adequate to excellent test re-test correlations between forms A and B, supporting the equivalency of the parallel versions, and an excellent inter-rater reliability of .95 for the Figure Copy subtest (the subtest with the greatest degree of subjectivity in scoring procedures). Convergent and divergent validity is supported by large ($\geq .5$: Cohen, 1988) correlations with other theoretically similar neuropsychological tests and lower or no correlations with theoretically dissimilar neuropsychological tests and unrelated measures. Analysis of variance (ANOVA) demonstrated that the SPANS indices differentiated between patients with ABI less than one year previously, patients with long-term neurological conditions, and healthy controls. Finally, the LAI and VPI evidenced capacity to differentiate between left and right hemisphere damage.

WTAR

The WTAR (The Psychological Corporation, 2001) is a reading test that enables prediction of pre-morbid intellectual functioning. It is composed of a list of 50 words with irregular pronunciations and has been shown to be valid for use with a traumatic brain injury population (Green et al., 2008). It was employed in order to include only healthy control participants with an estimated average IQ.

Design and procedure

This was a retrospective study applying psychometric methods to archival data collected by Burgess (2013) in order to evaluate the reliability and validity of the SPANS. A combination of descriptive, correlational and factor analytic designs were employed in order to address the different aims and hypotheses of the study.

During test development, the SPANS was administered as part of routine clinical practice by trained clinicians following the manual's standardised protocol. Efforts were made to keep distractions and pauses to a minimum. The SPANS was always the first test administered, in some cases followed by the WTAR. Standardised procedures were also followed for healthy controls, with randomised allocation of form. Further details can be seen in the test manual (Burgess, 2013). All data were entered in an anonymous format into an IBM SPSS Statistics database, which was then accessed by the author of the current study.

Quality assurance checks

The database was thoroughly scrutinised by the author to check for errors, which when found were resolved by referring to the original paper documentation. All analyses were checked by a statistician.

Ethical considerations

Ethical approval for the original data collection was granted by the Research Ethics Committee (see Appendix D). This specified that data could be accessed by other researchers with the permission of the original research lead, Dr Burgess. The University Research Governance Manager was fully informed of the protocol for the current study (Appendix E). All data accessed by the author were fully anonymised. Participants were given a website address where they could access information about the study and the results (Appendix F).

Data analysis

Descriptive statistics of means, standard deviations and ranges for all subtests/indices according to group and form is provided in Appendix G. Regarding reliability and validity analyses, Cronbach's α was used to assess the internal consistency of six of the SPANS index scales for each participant group and form. Spearman-Brown Prediction formula was used to assess reliability of the CFI as this is recommended for scales containing only two subtests (Eisinga, Grotenhuis, & Pelzer, 2012). To evaluate the discriminative validity, receiver operating characteristic (ROC) curves and classification accuracy statistics were calculated. Construct validity was assessed through exploratory factor analysis (EFA). Further details of these are provided in the relevant results section.

Results:

1) Reliability: Internal consistency

The internal consistency of each of the index scores was assessed by calculating Cronbach's α coefficients (or Spearman-Brown formula for CFI) for each participant group separately, a combined clinical group, and for the whole sample. Calculations were also made for each form separately and combined. The results are presented in Table 3 below, which also includes the results reported in the SPANS test manual for comparison. George and Mallery (2003) suggested the following guidelines for interpreting Cronbach's α: "≥.9 – Excellent, ≥.8 – Good, ≥.7 – Acceptable, ≥.6 – Questionable, ≥.5 – Poor, and <.5 – Unacceptable" (p. 231).

Table 3: Internal consistency of the SPANS index scores by group and form

Index	ABI			LTNC			Controls		
Form	Both	A	B	Both	A	B	Both	A	B
	n=61	n=49	n=12	n=35	n=28	n=7	n=122	n=61	n=61
ORI	.766	.761	.773	.771	.780	.720	.282	-.075	.326
ACI	.824	.792	.851	.783	.793	.731	.423	.476	.277
MLI	.894	.893	.835	.827	.760	.920	.669	.609	.744
LAI	.865	.851	.827	.735	.618	.902	.429	.314	.481
VPI	.875	.860	.879	.813	.826	.749	.391	.093	.574
PSI	.834	.768	.886	.859	.875	.778	.314	.248	.378
CFI	.752	.792	.582	.550	.542	.809	.497	.284	.624

Index	ABI & LTNC Combined			Whole sample			Results from
Form	Both	A	B	A	B	Both	Burgess (2013)
	n= 96	n= 77	n= 19	n= 138	n= 80	N= 218	N = 258
ORI	.769	.770	.754	.803	.807	.804	.785
ACI	.825	.796	.868	.801	.871	.831	.790
MLI	.897	.887	.895	.876	.929	.899	.899
LAI	.856	.829	.861	.839	.894	.866	.858
VPI	.868	.856	.873	.827	.871	.845	.851
PSI	.849	.814	.896	.808	.880	.838	.840
CFI	.757	.750	.745	.778	.821	.796	.728

As Table 3 shows, internal consistency coefficients were all adequate to excellent for the sample as a whole, regardless of form. The MLI, LAI and VPI demonstrated the highest α's, whilst CFI and ORI showed the lowest α's (but still at or approaching excellence). Looking at the clinical groups separately, internal consistencies were all adequate to excellent within the ABI group with the exception of the CFI form B. Within the LTNC group, all α's were adequate to excellent with the exception of CFI form A/forms combined, which was poor, and LAI form A, which was questionable. However, within the healthy control group internal consistencies are poor for all bar the MLI.

2) Discriminant validity: ROC curve analysis and classification statistics

Neurological condition vs. healthy controls

The ability of the SPANS to discriminate between healthy control participants and those with a neurological condition (ABI and LTNC groups combined) was examined through ROC analysis and calculation of classification statistics for each form separately and combined. ROC curves provide a complete measure of accuracy by plotting discriminative ability (true positive rate by false positive rate) across the whole spectrum of potential cut-offs (Kumar & Indrayan, 2011). Initially, ROC curves were calculated separately for each index scale and their corresponding subtests to establish whether the index scores were a better discriminator than the individual subtests. Only letter-number coding (LNC) outperformed its corresponding index scales (VPI and PSI) therefore only indices and this subtest are reported from hereon in. The ROC curves can be seen in Figures 1 - 3 and information relating to area under the curve (AUC) is in Table 4. AUC provides an overall measure of discrimination, with a score of one (1.0) representing perfect discrimination. Hosmer and Lemeshow (2000) suggest an AUC of ≥.70 is acceptable and ≥.80 is excellent.

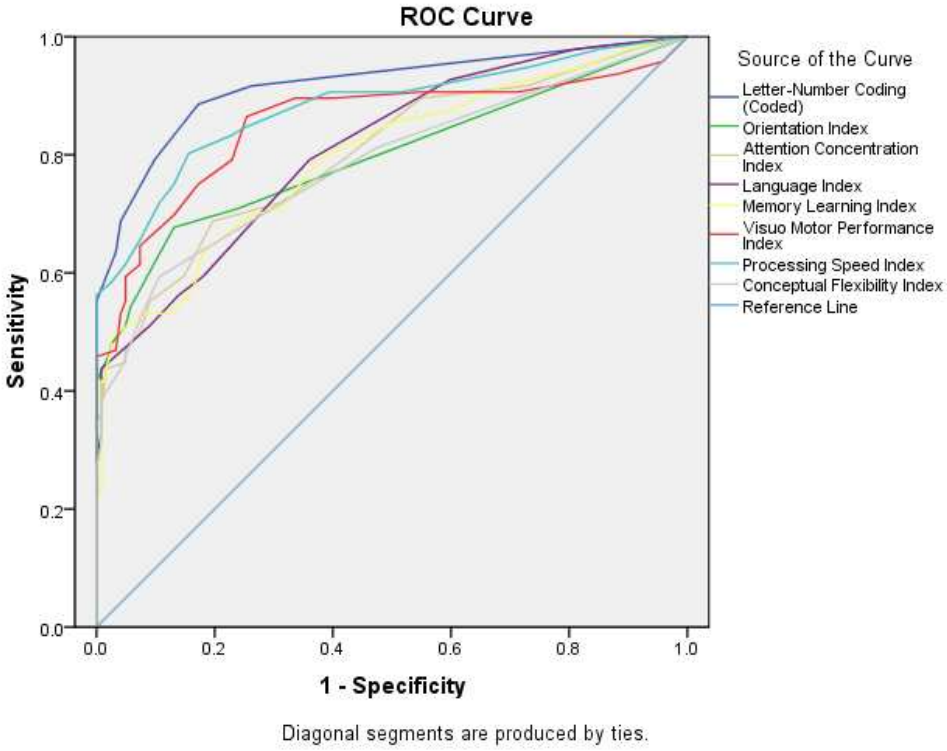

Figure 1: ROC curve for SPANS indices and LNC subtest for forms combined

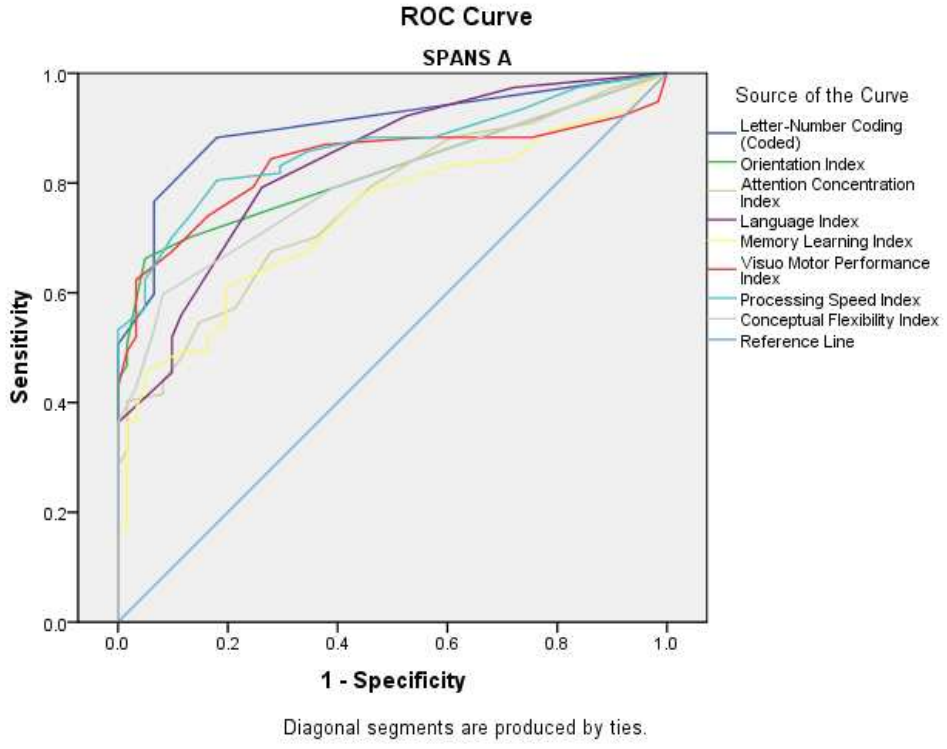

Figure 2: ROC curve for SPANS indices and LNC subtest for form A

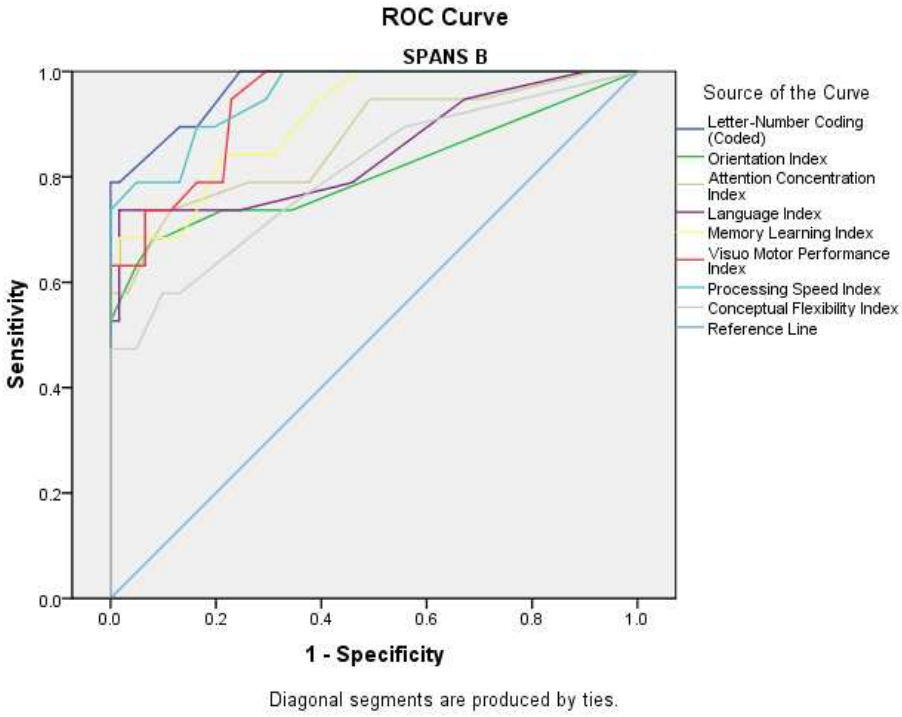

Figure 3: ROC curve for SPANS indices and LNC subtest for form B

As can be seen in Figures 1 - 3, the curves of all scales are above the diagonal 'line of no information' suggesting using the scale is better than guessing. LNC is the closest to the top left corner 'perfect' axis. This information is supported quantitatively (as shown in Table 4) by highly significant ($p<.001$) AUCs most of which approach or exceed excellence.

Table 4: Area under the curve statistics for SPANS indices and LNC subtest

Index/ Subtest	Form	AUC	Std. Error	Significance	95% Confidence Interval Lower Bound	Upper Bound
LNC	Both	.917	.021	.000	.877	.958
	A	.902	.027	.000	.850	.955
	B	.971	.017	.000	.925	1.000
ORI	Both	.797	.033	.000	.733	.861
	A	.822	.036	.000	.751	.892
	B	.809	.072	.000	.668	.950
ACI	Both	.801	.031	.000	.741	.862
	A	.763	.040	.000	.685	.841
	B	.868	.054	.000	.762	.973
LAI	Both	.810	.029	.000	.753	.867
	A	.835	.034	.000	.769	.901
	B	.847	.062	.000	.727	.968
MLI	Both	.797	.031	.000	.736	.858
	A	.739	.042	.000	.657	.822
	B	.912	.037	.000	.840	.983
VPI	Both	.853	.029	.000	.796	.909
	A	.836	.036	.000	.765	.907
	B	.937	.027	.000	.884	.990
PSI	Both	.881	.025	.000	.832	.930
	A	.862	.032	.000	.800	.924
	B	.954	.024	.000	.906	1.000
CFI	Both	.785	.033	.000	.721	.849
	A	.798	.037	.000	.724	.871
	B	.804	.064	.000	.679	.929

Youden's Index (max: sensitivity + specificity – 1: Youden, 1950) was used to calculate the optimal cut-off score for each index according to statistical criteria. However, Lincoln, Nicholl and Flannaghan (2003) suggested that in clinical practice, adequate diagnostic measures should have a sensitivity and specificity greater than 80% and 60% respectively. Therefore cut-offs were chosen that had the highest Youden's Index and were closest to Lincoln and colleagues' (2003) criteria. Following this classification statistics were calculated. Sensitivity (or true positive rate) refers to the proportion of people with an ABI correctly identified as such. Specificity (or true negative rate) refers to the proportion of healthy controls correctly identified as such. Likelihood ratio positive and negative (LR±) refers to the extent to which a test result changes the probability that a condition exists, or put another way how many times more likely a person with an ABI is to have a positive or negative test result compared to healthy controls. Positive predictive value (PPV) refers to the proportion of positive test results that are true positives, whilst negative predictive value (NPV) refers to the proportion of negative test results that are true negatives. This information is shown in Table 5.

Table 5: Classification statistics for SPANS indices and LNC subtest

	Form	Cut-off (Impaired if ≤)	Sens	Spec	LR+	LR-	PPV(%)	NPV(%)
LNC	Both	10	.89	.83	5.24	0.13	80	91
	A	10	.88	.82	4.89	0.15	86	84
	B	9	.90	.87	6.92	0.11	69	96
ORI	Both	21	.71	.76	2.96	0.38	70	77
	A	21	.70	.87	5.38	0.31	87	69
	B	21	.74	.79	3.52	0.33	53	91
ACI	Both	41	.72	.69	2.32	0.35	65	76
	A	41	.70	.64	1.94	0.47	71	63
	B	41	.79	.74	3.04	0.28	49	92
LAI	Both	50	.79	.64	2.19	0.33	63	80
	A	50	.79	.74	3.04	0.28	79	73
	B	46	.74	.98	3.70	0.27	92	92
MLI	Both	60	.79	.62	2.08	0.34	62	79
	A	60	.78	.56	1.77	0.39	69	67
	B	58	.84	.79	4	0.20	56	94
VPI	Both	63	.87	.75	3.48	0.17	73	88
	A	63	.84	.72	3	0.22	79	78
	B	63	.95	.77	4.13	0.06	57	98
PSI	Both	40	.80	.84	5	0.24	80	84
	A	40	.81	.82	4.5	0.23	85	77
	B	41	.90	.84	5.63	0.12	64	96
CFI	Both	26	.70	.72	2.5	0.42	66	75
	A	26	.69	.77	3	0.40	79	21
	B	26	.74	.67	2.24	0.39	41	89

As can be seen in Table 5, optimum cut-off scores and classification statistics were largely similar across the two forms, though form B tended to demonstrate lower PPVs. The LNC, PSI and VPI appear to be the most discriminative measures, with sensitivities above .80 and specificities above .70. A person with an ABI is 3 - 5 times more likely to obtain a score below cut-off on these measures than a healthy individual. The LAI, CFI, ORI and ACI were less good discriminatory measures, but still demonstrated adequate sensitivities mostly above .70 and specificities mostly above .60. Form B tended to outperform form A in terms of discriminative ability.

Left vs. right hemisphere damage

ROC curves were employed to assess the ability of the SPANS LAI and VPI to discriminate between left (LH) and right (RH) hemisphere damage. It was not possible to evaluate the SPANS forms separately due to small sample sizes, therefore forms are combined for the following analyses. ROC curves can be seen in Figures 4 and 5 followed by AUC information and classification statistics in Tables 6 and 7.

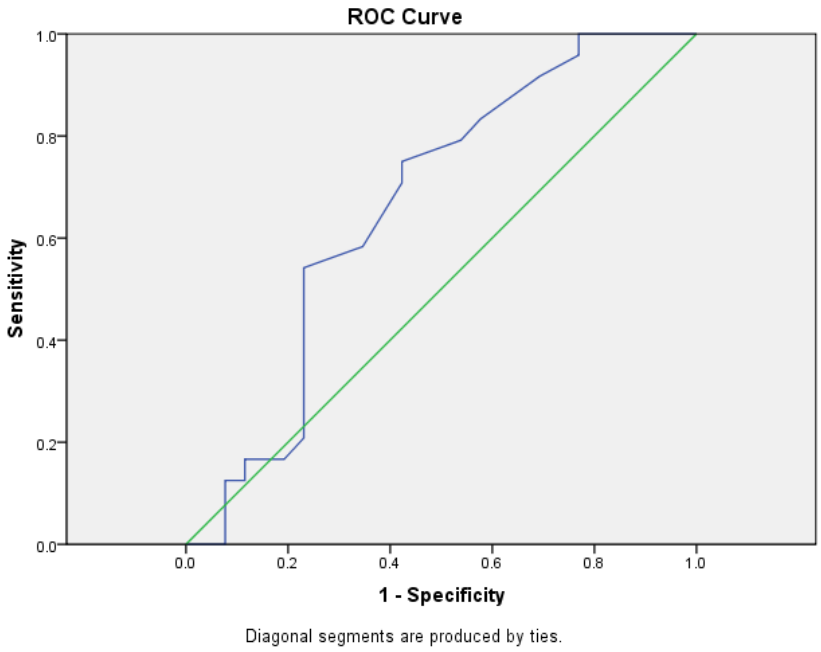

Figure 4: ROC curve showing VPI ability to detect RH damage

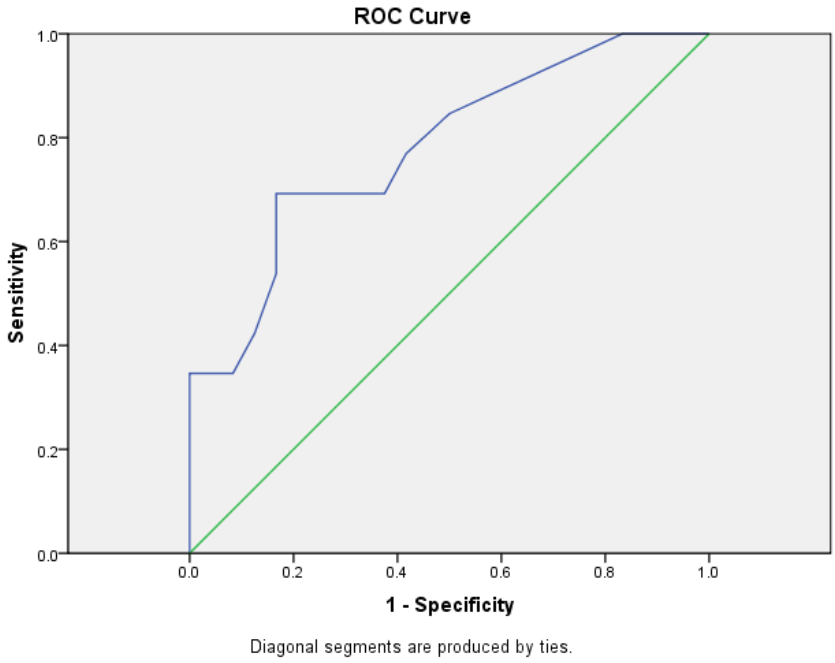

Figure 5: ROC curve showing LAI ability to detect LH damage

The ROC curves show that both VPI and LAI are able to indicate lateralisation above chance level, though LAI appears to be a better indicator. Both LAI and VPI significantly ($p<.05$) discriminate between left and right hemisphere damage and the AUC was acceptable for LAI and approaching the acceptable range for VPI.

Table 6: AUC information for VPI and LAI

				95% Confidence Interval	
	AUC	Std. Error	Significance	Lower Bound	Upper Bound
VPI	.663	.078	.048	.510	.817
LAI	.781	.064	.001	.655	.907

Table 7: Classification statistics for VPI and LAI

	Cut-off (if ≤)	Sens	Spec	LR+	LR-	PPV(%)	NPV(%)
VPI	54	.75	.58	1.79	0.43	62	72
LAI	42	.69	.83	4.06	0.37	81	71

Table 7 shows the optimal cut-offs for detecting lateralisation. The VPI approached acceptable sensitivity and specificity. A person with right hemisphere damage is almost twice as likely to score below the cut-off score on the VPI. The LAI demonstrated good specificity but lower sensitivity. However, the LR+ shows that a person with left hemisphere damage is over four times more likely to score below the cut-off score on the LAI. A score below cut-

off on the VPI indicates a 62% chance of right hemisphere damage, whilst a score below cut-off on the LAI indicates an 81% chance of having left hemisphere damage.

3) Exploratory Factor Analysis (EFA)

Factor analysis is a technique which reduces a large set of variables into a smaller set of variables seen to represent the underlying latent factors within a dataset. EFA was run on the entire sample and forms combined (N = 218), as theoretically the underlying constructs should not differ between clinical and control groups or between forms. Principal axis factoring was chosen over principal component analysis as the former method is favoured when there are pre-existing theoretical grounds for particular underlying constructs, whereas principal component analysis provides a descriptive empirical summary of a dataset (Tabachnick & Fidell, 2001). Principal axis factoring is also argued to be more accurate (Cliff, 1987) and to generalise better to confirmatory factor analysis (Floyd & Widaman, 1995). This is because it excludes error and unique variance and analyses covariance (communalities) only (Tabachnick & Fidell, 2001). Oblique rotation was chosen as this is the preferred method when theoretically one would expect factors to be related (Field, 2009).

Comrey and Lee (1992) suggest that a sample size of 200 is fair and the results of the Kaiser-Meyer-Olkin test, a statistical measure of the adequacy of sample size, was considered "superb" at .935 (Hutcheson & Sofroniou, 1999; cited in Field,2009). Bartlett's test of sphericity, a measure of whether overall correlations are large enough to indicate underlying factors, was also significant: χ^2 (528) = 5684.12, p<.001. Visual screening of the correlation matrix for all items showed that no items correlated at or above .90, implying no concerns

regarding multicollinearity or singularity (overly high or perfect correlations between variables, implying redundancy of items) (Field, 2009). Therefore EFA was deemed appropriate.

An initial unconstrained principal axis factoring indicated five factors with initial Eigenvalues greater than 1 (Kaiser's criterion: Kaiser, 1960). The scree plot was slightly ambiguous, with points of inflexion that would justify retaining both three and five factors. The initial five-factor solution was examined for items with a loading ≥ .4 (as recommended by Stevens, 2002). Items crossloading onto two factors at .32 or higher and with less than .2 magnitude difference were deleted (as recommended by Tabachnick & Fidell, 2001). This resulted in the deletion of nine items and left one factor composed of only two items, generally deemed "weak and unstable" (Costello & Osborne, 2005, p.5).

Next, an analysis was run forcing four factors to be retained. Following the above criteria, three items were deleted due to cross-loadings and two factors were left with only three items. Finally, an analysis was run forcing three factors to be retained, due to examination of the scree plot indicating that this was another point of inflexion. This resulted in eight items being deleted due to cross-loadings or loadings below .4: Orientation to Time and Place, Figures Recognition, Sustained and Divided Listening I and II, Counting Backwards, Monetary Calculations, and List Learning. This three-factor solution was retained as the resulting factors each had more than five strongly loading (>.5) items, indicating a solid factor (Costello & Osborne, 2005). The factors were interpretable theoretically, and together they accounted for 56.23% of the total variance, deemed reasonable (Streiner, 1994). Table 8 shows the factor loadings after rotation.

Table 8: Summary of factor loadings after rotation for final 3-factor solution

	Factor		
	1	2	3
Facial Expressions	**.668**		
Spatial Decision	**.631**		
3 and 1 Concept Test	**.624**		
Time Estimation	**.600**		
Figures Copy	**.593**		
Orientation to Time	.535		-.386
Object Recognition	**.531**		
Letter-Number Coding	**.514**	.303	
Unusual Views	**.479**		
Figures Recognition	.478		-.422
Sustained and Divided Listening I	.396	.308	
Orientation to Place	.307		
Following Directions		**.782**	
Yes/No Questions		**.774**	
Naming		**.730**	
Reading		**.719**	
Writing Sentences		**.671**	
Repetition		**.652**	
Similarities		**.616**	
Digit Span Forward		**.513**	
Counting Backwards	.362	.495	
Digit Span Backward		**.462**	
Orientation to Condition		**.421**	
Monetary Calculations	.372	.396	
Sustained and Divided Listening II	.326	.361	
List Recall			**-.874**
Object Recall			**-.845**
List Recognition			**-.794**
Figures Recall			**-.719**
Word-Symbol Paired Associates			**-.671**
Orientation to Prime Minister/President			**-.522**
List Learning		.461	-.509
Orientation to Person			**-.505**

Extraction Method: Principal Axis Factoring. Rotation Method: Oblimin with Kaiser Normalization. Loadings < .3 are suppressed. **Bold loadings represent items to be included in factor.**

The final EFA solution resulted in 25 subtests making up three factors which account for 56% of the total variance. The first factor appeared to relate predominantly to visual-motor performance and accounted for 47% of the variance. The second factor appeared to relate predominantly to language abilities and accounted for 5% of the variance. The third factor appeared to relate to memory and accounted for 4% of the variance. Examination of the factor correlation matrix showed moderate to high correlations between all factors, with factor three showing the highest correlations with factor one ($r = .59$) and factor two ($r = .60$).

Internal reliability of factors

Cronbach's α were calculated to check the internal consistency and reliability of the three factors determined through EFA. All factors demonstrated good reliability with Cronbach's $\alpha \geq .8$ (range: .836 - .886). All items demonstrated high item-to-total correlations and contributed to the factor, therefore no further items needed to be deleted.

Discussion

Reliability and validity of the SPANS

The current study provides further evidence of the validity of the SPANS for the assessment of cognitive function in people with an ABI. In support of the first hypothesis, internal reliability was shown to be mostly adequate to excellent within the clinical groups both separately and combined for all indices, irrespective of form. In keeping with the second hypothesis ROC curves showed that each index and the LNC subtest (irrespective of form) were able to significantly discriminate between healthy controls and patients with a

neurological condition, with all AUCs at acceptable or excellent levels. Furthermore, in keeping with previous studies (e.g. Capruso & Levin, 1992) classification statistics showed that the indices with more 'ability-focused' tests (Larrabee, 2008)such as LNC, VPI and PSI were the best discriminators, with all sensitivities and specificities above Lincoln and colleagues' (2003) criteria. The LAI and VPI demonstrated significant discriminative ability for lateralisation. Scores below optimal cut-offs on the LAI and VPI was indicative of left and right hemisphere damage respectively, in keeping with previous evidence of lateralisation of language and visual skills (Kolb & Whishaw, 2003). Regarding the third aim and hypothesis of the study, partial support for the construct validity of the SPANS was demonstrated by EFA which resulted in three theoretically meaningful factors composed of 25 subtests explaining 56% of the total variance, all with good internal reliabilities.

Within the healthy control group, Cronbach's α were predominantly poor. Examination of frequency distributions and subtest inter-correlations within the healthy control sample showed that most items were negatively skewed, had limited or zero variance and few significant strong correlations, which all tend to deflate Cronbach's α (Cortina, 1993). Other authors have similarly reported difficulties calculating internal consistency for healthy subjects due to limited variability (e.g. Bullard et al., 2004). The SPANS was designed to be challenging but passable by healthy subjects (Burgess, 2013), hence the negative skew and limited variability. Alternatively, the low internal consistencies could mean that the index scores are not tapping into distinct cognitive constructs and that therefore the test is measuring something different in healthy participants compared to clinical. Research has shown that using cognitive strategies (e.g. visualisation) can impact test performance (Ball et al., 2002). It is plausible that healthy participants, with their intact cognitive abilities, employed a broader range of strategies to solve test problems and that therefore each test item

may be measuring abilities other than what they purport to. However, it seems more likely that the results are a spurious effect of the limited variance. As the SPANS was designed to identify cognitive impairment in people with an ABI rather than to assess normal variability in cognitive skills in the general population, it is more important that it demonstrates good reliability within the clinical groups.

The ROC curves and classification statistics offer initial support for the discriminative validity of the SPANS and provide clinicians with information regarding its diagnostic accuracy. Though performance of the two forms was largely similar, there were discrepancies between forms on the LAI and MLI such that form B demonstrated better specificity and LR+ compared to form A. This begs the question of whether the parallel versions are truly equivalent, or whether this is a spurious effect of sampling differences. It is difficult to attribute any observed differences between forms on the LAI as being due to non-equivalence as this index has only one subtest which varies between forms (Naming). Therefore sampling differences may be responsible. A smaller proportion of participants administered form B had left hemisphere damage and therefore may have had fewer focal language deficits. Regarding the MLI, all subtests vary between forms and therefore the observed differences in classification accuracy may reflect real differences between the tests. This may be the case despite the great care that was taken to ensure equivalence (see Burgess, 2013 for full details) during the design and development stages of the SPANS. However, sampling differences again cannot be ruled out.

The reported structure of the SPANS is five distinct and two overlapping index scales (Burgess, 2013). This structure was only partially supported by the results of the EFA in the

current study. The most interpretable and strongest solution indicated three factors: language, visual-performance and memory/learning. These factors corresponded largely to LAI, VPI and MLI indices, with a few differences. Specifically, the visual-performance factor excludes the Figure Recognition subtest as this loaded equally well onto the memory factor. Theoretically this makes sense as the task clearly involves both visual processing and encoding. The visual-performance factor also included the Time Estimation subtest. Other studies have shown impairments in time estimation to be linked to damage in temporo-parietal regions of the brain (Barabassy, Beinhoff, & Riepe, 2007), which are also strongly associated with visuospatial processing and knowledge of numbers (Kosslyn, 2007; Blakemore & Frith, 2005). All subtests from the LAI demonstrated the strongest loadings onto the language factor. Weaker loadings came from Digit Span and Orientation to Condition subtests, which may be due to the reliance on language in order to understand the instructions/question and produce a verbal answer with adequate sophistication. Similarly, the strongest loadings onto the memory factor came from the original MLI subtests, with additional weaker loadings from the Orientation to Prime Minister/President and Person subtests. It could be argued that these both rely on retrieval from long-term memory systems.

Factor analytic studies of other neuropsychological assessment batteries have similarly not always supported the reported test structure. As mentioned in the introduction, a principal components analysis of the RBANS resulted in a two-factor solution, as opposed to the five factor test structure purported by the test (Wilde, 2006). Cognitive constructs of orientation, attention, concentration, processing speed and executive function appear to be the hardest to assess in isolation, perhaps as these are more shared cognitive processes on which successful completion of any task will rely. For example, a person's ability to attend to task will clearly affect performance in all domains (Hodges, 1999). Furthermore as noted in the introduction

attention is not a unitary concept (Sohlberg & Mateer, 1989), nor are executive functions (Poulin, Korner-Bitensky, & Dawson, 2013). It is therefore perhaps not surprising that these do not emerge as independent factors in EFA.

The SPANS is one of only a few brief neuropsychological assessment batteries that has been designed specifically for the purpose of assessing cognitive function in patients with an ABI. It was developed following extensive review of the theoretical and empirical literature, as summarised in the introduction. In comparison to similar existing batteries (RBANS and NAB-SM), the SPANS demonstrated better internal reliabilities across all its index scores for the clinical group (see McKay et al., 2007; Zgaljardic & Temple, 2010). Classification statistics are not available for the NAB-SM, and the SPANS indices have higher specificities (though generally lower specificities) compared to the RBANS (see McKay et al., 2008). Similarly, there have not been any factor analytic studies of the NAB-SM, and as mentioned a principal component analysis of the RBANS did not support its five-factor structure (Wilde, 2006). Therefore, the current study provides preliminary evidence for the design aim of the SPANS in addressing the psychometric limitations of existing brief neuropsychological batteries and appears to be a reliable and valid tool for the assessment of cognitive function in people with an ABI.

Limitations of the study

There are a number of limitations to the current study which impact on the conclusions that can be drawn. Due to the archival and retrospective nature of the study, there was a relative lack of control over sampling design and composition of participants. Participants were

recruited from a relatively homogeneous area geographically and so may not generalise to other regions. Sampling was non-consecutive introducing the possibility of bias in the data (e.g. the absence of more severely impaired participants). The data from participants who were unable to fully complete the SPANS was excluded (as is the case with most neuropsychological batteries), meaning the current sample is not representative of all people presenting to neurorehabilitation services and may reflect those with relatively less severe injuries. The clinical group is heterogeneous in terms of onset, cause, severity and location of injury. Sample sizes are relatively small, particularly for SPANS form B and the lack of systematic control over sampling procedures meant that groups were not necessarily comparable on all variables. The above factors all limit the generalizability of the results, a common issue in brain injury research (Car-Blanchard, 2004).

Although establishing the equivalency of the parallel versions of the SPANS was not a specific aim of the current study, attempts were made to compare the forms throughout the analyses to improve control. This was not always possible or valid due to sample sizes (e.g. in evaluating discriminative validity for lateralisation) and interpretations of differences are complicated by the lack of control over sampling procedures and therefore potential differences between groups. For the EFA, the entire sample was analysed as a whole in order to have a large enough sample size, but some authors argue that underlying factor structure needs to be determined in the group for which the test is designed independently, as sample composition can impact on the results (Delis, Jacobsen, Bondi, Hamilton, & Salmon, 2003).

Clinical/theoretical implications

Neuropsychological assessment has a critical role to play in brain injury rehabilitation in terms of assessing a patient's cognitive strengths and weaknesses, learning potential, capacity to make complex decisions, and predict care needs and functional outcomes (Harvey, 2012). There has been a gap for clinicians for brief but comprehensive cognitive assessments with strong theoretical and empirical underpinning to their development and sound psychometric properties. The current study provides preliminary support for the clinical utility, reliability and validity of the SPANS for the assessment of cognitive impairment in patients with an ABI or other neurological conditions. The study also strengthens previous research that has shown 'ability-focused' tests to discriminate between healthy controls and patients with a neurological condition better than semantic knowledge tests (e.g. Larrabee, 2008).

Future research directions

In order to provide further validation of the clinical utility, reliability and validity of the SPANS and to determine generalizability of the current results, it would be helpful to repeat the current analyses in an independent sample. Further research assessing the predictive validity of the SPANS, e.g. of future cognitive status, return to work, or community integration would also be beneficial. In order to ascertain the equivalence of the parallel forms, an experimental design could be employed using either repeated measures over a fixed time period in a counterbalanced order, or using a carefully matched independent groups design. This would help to calculate potential practice effects in order to be able measure change due to treatment effects more precisely. With a larger, independent sample confirmatory factor analysis could be conducted to test the seven or five factor structure of the test, or the three factor structure produced in the current study.

Although the SPANS was designed primarily for use with working age ABI populations, it has the potential to expand into other groups, such as children and adolescents, older adults, other clinical conditions such as dementia, and other settings such as forensic. Further research is needed to assess its suitability and psychometric properties in these areas.

Conclusions

Assessing cognition is an important role for clinical psychologists in neurorehabilitation services but there is a lack of theoretically and psychometrically valid brief neuropsychological assessment batteries available. The SPANS was designed to meet this need. The results of the current study demonstrated that the SPANS had high internal reliability when administered to patients with an ABI or other neurological condition. It also demonstrated good discriminative ability between healthy and clinical patients and detected left or right hemisphere damage with some confidence. The visual performance, language and memory indices were strongly supported by findings from EFA, suggesting that these are valid and reliable constructs and map to some degree the a priori structure of the measure. Overall, the SPANS appears to at least match and even outperform similar existing measures and meet its design aims of being a reliable and valid tool for the assessment of cognitive impairment in patients with an ABI.

References

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., ... & Willis, S. L. (2002). Effects of cognitive training interventions with older adults. JAMA: the journal of the American Medical Association, 288(18), 2271-2281.

Barabassy, A., Beinhoff, U., & Riepe, M. W. (2007). Cognitive estimation in mild Alzheimer's disease. Journal of Neural Transmission, 114(11), 1479-1484.

Bigler, E. D. (2007). A motion to exclude and the 'fixed' vs. 'flexible' battery in 'forensic' neuropsychology: Challenges to the practice of clinical neuropsychology. Archives of Clinical Neuropsychology, 22, 45–51.

Blakemore, S. J., & Frith, U. (2005). The learning brain: Lessons for education. Malden MA: Blackwell publishing.

Bullard, S. E., Fein, D., Gleeson, M. K., Tischer, N., Mapou, R. L., & Kaplan, E. (2004). The Biber cognitive estimation test. Archives of clinical neuropsychology, 19(6), 835-846.

Burgess, G.H. (in press). Short Parallel Assessments of Neuropsychological Status (SPANS). Oxford, UK: Hogrefe, Ltd. Capruso, D. X., & Levin, H. S. (1992). Cognitive impairment following closed head injury. Neurologic clinics, 10(4), 879-893.

Car-Blanchard, M. (2004). Research Findings from the Traumatic Brain Injury Model Systems.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. Neuropsychology review, 13(4), 181-197.

Chelune, G. J. (2010). Evidence-based research and practice in clinical neuropsychology. The Clinical Neuropsychologist, 24(3), 454-467.

Cliff, N. (1987). Analyzing multivariate data. San Diego, CA: Harcourt Brace Jovanovich.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. New York, NY: Routledge.

Comrey, A. L., & Lee, H. B. (1992). A first course in factor analysis. New York, NY: Routledge.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of applied psychology, 78(1), 98.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation, 10(7), 1-9. Retrieved from http://pareonline.net/pdf/v10n7.pdf

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. Journal of the International Neuropsychological Society, 9(06), 936-946.

Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. Neuropsychology, 9(1), 80.

Dikmen, S., McLean Jr, A., Temkin, N. R., & Wyler, A. R. (1986). Neuropsychologic outcome at one-month post injury. Archives of Physical Medicine Rehabilitation, 67(8), 507-513.

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2012). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. International journal of public health, 1-6.

Field, A. (2009). Discovering statistics using SPSS. London: Sage publications.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. Psychological assessment, 7(3), 286.

George, D., & Mallery, M. (2003). Using SPSS for Windows step by step: a simple guide and reference. Boston, MA: Allyn & Bacon

Green, R. E., Melo, B., Christensen, B., Ngo, L. A., Monette, G., & Bradbury, C. (2008). Measuring premorbid IQ in traumatic brain injury: An examination of the validity of the Wechsler Test of Adult Reading (WTAR). Journal of Clinical and Experimental Neuropsychology, 30(2), 163-172.

Hanks, R. A., Millis, S. R., Ricker, J. H., Giacino, J. T., Nakese-Richardson, R., Frol, A. B., ... & Gordon, W. A. (2008). The predictive validity of a brief inpatient neuropsychologic battery for persons with traumatic brain injury. Archives of physical medicine and rehabilitation, 89(5), 950-957.

Harvey, P. D. (2012). Clinical applications of neuropsychological assessment. Dialogues in clinical neuroscience, 14(1), 91.

Headway: the brain injury association. About brain injury. Retrieved from https://www.headway.org.uk/About-Brain-Injury.aspx

Hodges, J. R. (1999). Cognitive assessment for clinicians. Oxford, NY: Oxford University Press.

Holdnack, H.A. (2001). Wechsler Test of Adult Reading: WTAR. San Antonio. The Psychological Corporation

Hosmer, D. W., & Lemeshow, S. (2000). Model-Building Strategies and Methods for Logistic Regression. Applied Logistic Regression, Second Edition, 91-142.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and psychological measurement.

King, N. S., & Tyerman, A. (2003). Neuropsychological presentation and treatment of head injury and traumatic brain damage. In P. W. Halligan, U. Kischka, & C. Marshall (Eds.), Handbook of clinical neuropsychology, (pp. 487-505). Oxford, NY: Oxford University Press.

Kline, P. (1999). The handbook of psychological testing (2nd Ed.). London: Routledge

Kolb, B., & Whishaw, I. Q. (2003). Fundamentals of human neuropsychology. New York, NY: Worth Publishers.

Kosaka, B. (2006). Neuropsychological assessment in mild traumatic brain injury: A clinical overview. British Columbia Medical Journal, 48(9), 447.

Kosslyn, S. (2007). Cognitive Psychology: Mind and Brain. New Jersey: Prentice Hall

Kreutzer, J. S., Gordon, W. A., Rosenthal, M., & Marwitz, J. (1993). Neuropsychological characteristics of patients with brain injury: Preliminary findings from a multicenter investigation. The Journal of Head Trauma Rehabilitation, 8(2), 47-59.

Kumar, R., & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. Indian pediatrics, 48(4), 277-287.

Larrabee, G. J. (2008). Flexible vs. fixed batteries in forensic neuropsychological assessment: Reply to Bigler and Hom. Archives of Clinical Neuropsychology, 23(7), 763-776.

Larson, E. B., Kirschner, K., Bode, R., Heinemann, A., & Goodman, R. (2005). Construct and predictive validity of the Repeatable Battery for the Assessment of Neuropsychological Status in the evaluation of stroke patients. Journal of Clinical and Experimental Neuropsychology, 27(1), 16-32.

Leśniak, M., Bak, T., Czepiel, W., Seniów, J., & Członkowska, A. (2008). Frequency and prognostic value of cognitive disorders in stroke patients. Dementia and geriatric cognitive disorders, 26(4), 356-363.

Lincoln, N. B., Nicholl, C. R., Flannaghan, T., Leonard, M., & Van der Gucht, E. (2003). The validity of questionnaire measures for assessing depression after stroke. Clinical Rehabilitation, 17(8), 840-846.

McKay, C., Casey, J. E., Wertheimer, J., & Fichtenberg, N. L. (2007). Reliability and validity of the RBANS in a traumatic brain injured sample. Archives of Clinical Neuropsychology, 22(1), 91-98.

McKay, C., Wertheimer, J. C., Fichtenberg, N. L., & Casey, J. E. (2008). The repeatable battery for the assessment of neuropsychological status (RBANS): clinical utility in a traumatic brain injury sample. The Clinical Neuropsychologist, 22(2), 228-241.

Nabors, N. A., Millis, S. R., & Rosenthal, M. (1997). Use of the Neurobehavioral Cognitive Status Examination (Cognistat) in traumatic brain injury. The Journal of Head Trauma Rehabilitation, 12(3), 79-84.

Pachet, A. K. (2007). Construct validity of the Repeatable Battery of Neuropsychological Status (RBANS) with acquired brain injury patients. The Clinical Neuropsychologist, 21(2), 286-293.

Poulin, V., Korner-Bitensky, N., & Dawson, D. R. (2013). Stroke-specific executive function assessment: A literature review of performance-based tools. Australian occupational therapy journal, 60(1), 3-19.

Powell, J. B., Cripe, L. I., & Dodrill, C. B. (1991). Assessment of brain impairment with the Rey Auditory Verbal Learning Test: A comparison with other neuropsychological measures. Archives of Clinical Neuropsychology, 6, 241–249.

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

Randolph, C. (1997). Differentiating vascular dementia from alzheimer's disease: The role of neuropsychological testing. Clinical Geriatrics, 5, 77–86.

Russell, E. W., Russell, S. L., & Hill, B. D. (2005). The fundamental psychometric status of neuropsychological batteries. Archives of Clinical Neuropsychology, 20(6), 785-794.

Schoenberg, M. R. (2011). The Little Black Book of Neuropsychology: A Syndrome-based Approach. M. R. Schoenberg, & J. G. Scott (Eds.). Springer.

Sohlberg, M. M., & Mateer, C. A. (1989). Introduction to cognitive rehabilitation: Theory and practice. New York, NY: Guilford Press.

Stern, R. A., & White, T. (2003). Neuropsychological Assessment Battery. Lutz, FL: Psychological Assessment Resources.

Stevens, J. P. (2002). Applied multivariate statistics for the social sciences (4th ed.). Hillsdale, NJ: Erlbaum.

Streiner, D.L. (1994). Figuring out factors: the use and misuse of factor analysis. Canadian Journal of Psychiatry, 39, 135-140.

Sweet, J. J., Nelson, N. W., & Moberg, P. J. (2006). The TCN/AACN 2005 "salary survey": Professional practices, beliefs, and incomes of U.S. neuropsychologists. Clin. Neuropsychol., 20, 325–364.

Tabachnick, B. G., & Fidell, L. S. (2001). Using Multivariate Statistics. (4th ed.). Boston, Mass: Allyn & Bacon.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale. The Lancet, 304(7872), 81-84.

Temple, R. O., Zgaljardic, D. J., Abreu, B. C., Seale, G. S., Ostir, G. V., & Ottenbacher, K. J. (2009). Ecological validity of the neuropsychological assessment battery screening module in post-acute brain injury rehabilitation. Brain injury, 23(1), 45-50.

Turner-Stokes, L., Nair, A., Sedki, I., Disler, P. B., & Wade, D. T. (2011). Multi-disciplinary rehabilitation for brain injury in working-age adults, Cochrane Database Systematic Reviews, 3, 1-43.

Warrington, E. K., & James, M. (1991). The visual object and space perception battery. Bury St Edmunds: Thames Valley Test Company.

Wilde, M. C. (2006). The validity of the repeatable battery of neuropsychological status in acute stroke. The Clinical Neuropsychologist, 20(4), 702-715.

Williamson, D. J. G., Scott, J. G., & Adams, R. L. (1996). Traumatic brain injury. In R. L. Adams, O. A. Parsons, J. L. Culberston, & S. J. Nixon (Eds.). Neuropsychology for clinical

practice: Etiology, assessment, and treatment of common neurological disorders. (pp. 9-64). Washington, DC: American Psychological Association.

Youden, W. J. (1950). Index for rating diagnostic tests, Cancer, 3, 32-35.

Zgaljardic, D. J., & Temple, R. O. (2010). Reliability and validity of the Neuropsychological Assessment Battery-Screening Module (NAB-SM) in a sample of patients with moderate-to-severe acquired brain injury. Applied neuropsychology, 17(1), 27-36.

Zgaljardic, D. J., Yancy, S., Temple, R. O., Watford, M. F., & Miller, R. (2011). Ecological validity of the screening module and the Daily Living tests of the Neuropsychological Assessment Battery using the Mayo-Portland Adaptability Inventory-4 in postacute brain injury rehabilitation. Rehabilitation psychology, 56(4), 359.

Section C

Critical appraisal of the research process

Jennifer Attwood

July 2013

Word count:

1978

A thesis submitted in partial fulfilment of the requirements of

Canterbury Christ Church University for the degree of

Doctor of Clinical Psychology

What research skills have you learned and what research abilities have you developed from undertaking this project and what do you think you need to learn further?

The whole process of developing a research proposal through extensive reading of the wider literature, formulating research questions and hypotheses, exploring different potential research designs and methodologies and then submitting a proposal for peer review was a relatively novel experience for me, not having come into training from a research background. This part of the research process helped me to develop skills in concisely summarising existing theoretical and empirical literature in order to provide a clear rationale for the research aims and hypotheses, and also in describing and justifying research designs and methodology in a clear and explicable manner. Having my initial project turned down at the second ethics review really highlighted to me the importance of this stage of the research process and the necessity of being able to provide an account of a project that can be readily understood in lay terms and that has clear, positive, clinical implications. Though it felt very difficult at the time, this occurrence enabled me to reflect on the ethical issues with conducting basic/pure research compared to more applied research (Roll-Hansen, 2009), particularly when involving human participants. It also highlighted the importance of involving service users and other stakeholders in the design stages of research in order to ensure that research is acceptable to the people it most concerns. Thus for my second research proposal I took greater care in ensuring that the clinical implications were very apparent.

Conducting secondary analyses of existing datasets has some obvious examples in terms of removing the difficulties inherent with recruitment and data collection and possible confounds of experimenter bias (Shultz, Hoffman, & Reiter-Palmon, 2005). However, the challenges involved with this type of research were perhaps greater than I had originally anticipated. The majority of teaching received during training is tailored to gathering and

analysing new data, therefore there was little guidance available as to how to go about the process or what pitfalls to look out for. For example, following my review of the literature, I developed numerous potential research questions and designs, many of which turned out not to be possible with the existing dataset due to issues with the sampling procedures, sample size, or research design meaning that assumptions for certain statistical procedures could not be met. Whilst this was a time consuming and at times frustrating process, it also meant that I developed substantial knowledge about psychometric theory and methods, statistical analyses, and became much more proficient in exploring and managing datasets.

Receiver operating characteristic curves and classification statistics are a fundamental tool in clinical practice for evaluating the diagnostic or classification accuracy of a clinical measure (Zweig & Campbell, 1993). Hence gaining knowledge and understanding of and ability to apply these procedures has been an invaluable and challenging learning experience. I was able to be largely self-directive and autonomous in my approach to this learning, as this was not an area that my supervisors were familiar with themselves. This was useful in terms of increasing my confidence in learning and applying novel statistical procedures, but also in enabling me to feel a greater sense of ownership over the research project, despite not having collected the data myself. In order to conduct the exploratory factor analysis, I gratefully sought out consultation from a statistician. Having this expertise was essential to guide me on the numerous complex decisions required for conducting factor analysis, such as which type of factor analysis to use, how many factors to retain, which type of rotation to employ, how to interpret the factors, how to establish validity of the analysis and so forth. Having my decision-making process validated by someone with much greater knowledge and expertise helped me feel assured that the steps I had chosen were justified. In future I would feel more able to conduct this kind of analyses independently, and quantitative research more generally.

An area of research I feel I need further development and learning in would be in qualitative methods, as this is not something I was able to employ within the current study.

If you were able to do this project again, what would you do differently and why?

Ideally, given significantly more time, I would have liked to have been able to collect additional data in order to increase sample sizes but also to attempt to address some of the variability between-groups on some of the variables, particularly for those who were administered form B. Alternatively, amending the sampling procedures so that equal numbers of form A and form B were administered in a randomised but stratified manner, according to the key demographic and injury-related variables, would have enabled further analyses, e.g. of form equivalency, to be undertaken and strengthened the resulting conclusions. Collecting further data would also have given me the advantage of learning how to administer the test and expediting the process of familiarising myself with the test materials and the nature of each subtest, which was important for interpreting the results in light of the neuropsychological theory and research underpinning the choice of subtest. Unfortunately, due to the delay in commencing this project it was not possible to include further data collection with the proposal as this would have been subject to additional ethics review, as well as the additional time in recruitment.

An original aim of the project was to look at test form equivalency and quantify any potential practice effects using the repeated measures data. Unfortunately, following exploration of the dataset and gathering further information about the sampling procedures and study design, it became apparent that it would not be possible to draw any firm conclusions from the results. Feeding this back to my supervisor felt quite sensitive as I did not wish to appear critical or

unappreciative of the extensive amount of time and effort that had been spent on developing the measure and collecting the data; however I also did not wish to conduct analyses that I felt were not valid from a more stringent scientific/psychometric point of view. Again, adapting sampling procedures or collecting additional data could have resolved this issue. To enhance interpretation of the differences between the forms, I could have calculated confidence intervals for sensitivity and specificity figures. This is not something I had seen reported in other published studies and unfortunately time limitations did not allow me to do this.

As a consequence of doing this study, would you do anything differently in regards to making clinical recommendations or changing clinical practice and why?

The conclusions from the review of the extant literature clearly identified that there is a need in neuro-rehabilitation services for a brief but comprehensive, easy to administer, psychometrically reliable and valid neuropsychological tool for assessing cognitive function in people with an acquired brain injury. The current study provides preliminary support for the reliability, discriminative validity and factor structure of the Short Parallel Assessments of Neuropsychological Status (SPANS). Therefore, as a consequence of doing this study I would recommend the use of the SPANS in neuro-rehabilitation services in order to ascertain patients' level of cognitive functioning and track change over time. However, that said it would also be important for clinicians to continue to keep abreast of further research conducted with the SPANS as there is still a need for further validation studies.

Clinical psychology training and current models of health care emphasise the importance of being a scientist practitioner (Barlow, Hayes, & Nelson, 1984), part of which requires using reliable and valid assessment tools/outcome measures and promoting evidenced-based

practice. However, I had not fully considered the issues, from a psychometric theory perspective, of using an amalgam of different neuropsychological tests that have not been co-normed together. My experience of conducting neuropsychological assessments clinically both prior to and on placement, as well as my understanding from reading the literature (e.g. Kosaka, 2006), is that this is the most common approach taken by clinicians when carrying out neuropsychological assessments. As a result of completing this study, I would recommend that clinicians use a co-normed battery wherever possible. Where this is not possible, I would recommend that clinicians ensure that they are familiar with the normative data for the tests they are using and think about issues to do with comparability of norms, risk of false positive/negative results due to random error, and normal variation in abilities in the target population (for further critique of flexible batteries see Russell, Russell, & Hill, 2005).

My understanding of diagnostic/classification statistics (e.g. sensitivity and specificity) has enhanced as a result of doing this study and this has made me more aware of the risks and likelihood ratios of making false positive or false negative conclusions from tests. The importance of this will vary considerably depending on what a test is being used for and the consequences of having or not having a diagnosis. Arguably, within neurorehabilitation services, it may be better to falsely claim that someone has a cognitive impairment and allow them access to services and support rather than for cognitive impairment to go undetected and for a person to suffer the consequences in isolation. I will certainly pay more attention to this area when using standardised measures of any sort in clinical practice.

If you were to undertake further research is this area what would that research project seek to answer and how would you go about doing it?

A key area of research that appears to be relatively lacking in the broader field of neuropsychological assessment is evidence for how the results of neuropsychological assessment inform treatment planning and delivery and the impact of this on patient outcomes both short and long-term. This is a difficult area to study as it would not be possible to conduct a randomised controlled trial with some patients having a neuropsychological assessment and others not due to ethical issues of denying people an intervention predicted to be beneficial to their recovery (Temple & Ellenberg, 2000). An alternative possibility would be to compare two different measures, such as the SPANS and an equivalent test, on the results and recommendations of the assessment and subsequent patient outcomes, using an experimental or quasi-experimental design. A third possibility would be to conduct a retrospective study comparing neuro-rehabilitation patients who have or have not received a neuropsychological assessment in terms of their rehabilitation programs and outcomes. However, this approach would be limited by possible confounds between the samples as people who are referred for neuropsychological assessment are likely to have different presentations from those who are not.

In terms of further research relating more specifically to the SPANS, establishing the equivalence of the parallel versions of the test and quantifying potential practice effects appears essential, given the observed differences in performance in the current study, in order for clinicians to have confidence in the SPANS' ability to track recovery over time. To do this, I would recruit people with long-term and stable neurological conditions, in order to have enough variability in performance within the sample whilst not expecting further

recovery/decline in cognitive function. I would employ an experimental repeated measures design, administering form A and form B in a counterbalanced order at a set time interval apart, e.g. one week. Ideally, time between testing could be a moderating variable in order to see whether practice effects change time. Conducting research with people who have neurological conditions is challenging due to issues around prevalence, recruitment, informed consent, and physical or cognitive ability to participate. Therefore I acknowledge that to conduct this research would require extensive resources and time. To make it more feasible, it would be best if it could be conducted in a service that was already using the SPANS as part of standard clinical practice, in order to not add an additional burden onto a clinical team.

References

Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). The scientist practitioner: Research and accountability in clinical and educational settings (Vol. 128). New York, NY: Pergamon Press.

Kosaka, B. (2006). Neuropsychological assessment in mild traumatic brain injury: A clinical overview. British Columbia Medical Journal, 48(9), 447.

Roll-Hansen, N. (2009). Why the distinction between basic (theoretical) and applied (practical) research is important in the politics of science. London School of Economics and Political Science, Contingency and Dissent in Science Project.

Russell, E. W., Russell, S. L., & Hill, B. D. (2005). The fundamental psychometric status of neuropsychological batteries. Archives of Clinical Neuropsychology, 20(6), 785-794.

Shultz, K. S., Hoffman, C. C., & Reiter-Palmon, R. (2005). Using archival data for IO research: advantages, pitfalls, sources, and examples. The Industrial-Organizational Psychologist, 42(3), 31-37.

Temple, R., & Ellenberg, S. S. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Annals of Internal Medicine, 133(6), 455-463.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry, 39(4), 561-577.

Appendix A: Literature Search

All databases searched from their establishment to March 2013

Ovid Medline search strategy (MeSH terms in italics):

Brain Injuries or brain injuries or head injuries or Craniocerebral Trauma or Brain Ischemia or Stroke or Brain Damage, Chronic or Hypoxia-Ischemia, Brain or brain damage
AND
Neuropsychological Tests or neuropsychological assessment or cognitive assessment or screening tests
AND
Psychometrics or psychometrics or Reproducibility of Results or reliability or validity
LIMIT: humans and English language.

Result: 1274 articles

PsychINFO search strategy (subject headings in italics):

Brain injuries or Traumatic Brain Injuries or head injuries or Head Injuries or brain damage or Brain damage or Cognitive Impairment or Cerebrovascular Accidents
AND
Neuropsychological assessment or Neuropsychological Assessment or cognitive assessment or Cognitive Assessment or screening tests or Screening Tests
AND
Psychometrics or psychometrics or Test Validity or validity or Test Reliability or reliability
LIMIT: humans, English language and peer reviewed journals.

Result: 980 articles

EBM Reviews search strategy.
Neuropsychological assessment or cognitive assessment or screening tests
AND
Psychometrics or reliability or validity
AND
Brain injury or head injury or brain damage.

Result: 4 articles – none relevant.

Total elicited studies: 2,258

This systematic search was supplemented by an unsystematic search on Google using individual test names and reference to Lezak (2004) in order to gather further information about the tests as needed.

Neuropsychological tests were included if: 1) they were designed to assess for cognitive impairment in multiple domains and separate scores for each domain could be calculated; 2) were performance-based measures given direct to patients; 3) could be administered in one hour or less; and 4) had been published in English. Studies involving individual standardised tests, flexible batteries, standardised batteries with administration times of over one hour, tests which only give a global score, or unpublished tests (e.g. those in development that were not accessible) were excluded.

Appendix A: Literature Search

Papers relating to each neuropsychological test were included if: a) the main aim of the study was to investigate the psychometric properties (e.g. reliability or validity) of the test; and b) participants were predominantly working age (16 – 65 years old), English-speaking, adults with an ABI. Studies including older adults were retained if the mean age was less than 65 years old. Studies conducted solely with children and adolescents or older adults were excluded, as were studies with neurodevelopmental (e.g. learning disabilities) or neurodegenerative (e.g. dementia syndromes, Parkinson's disease, Multiple Sclerosis) conditions. Physical illnesses with associated cognitive impairments (e.g. HIV) were also excluded.

The initial search resulted in 2,258 hits across the databases. Titles and abstracts were screened according to the above criteria. In some cases where the required information was not clearly stated (e.g. no ages/diagnosis/name of test given), or apparent from the name of the journal (e.g. Dementia) the full text was sought. None of these met inclusion criteria. Removal of duplicates resulted in 15 papers meeting the inclusion and exclusion criteria. Searching the reference lists did not add any papers; nor did a second search on PsychINFO using the individual test names combined with "psychometrics".

Appendix B Summary of included studies

Ref	Aim(s) & Measure(s)	Participants	Analysis & Results
Doninger, N. A>, Bode, R. K., Heinemann, A. W., & Ambrose, C. (2000). Rating scale analysis of the neurobehavioural cognitive status examination. *Journal of Head Trauma Rehabilitation, 15(1),* 683-695.	Evaluate extent **NCSE** items represent a unidimensional construct. Investigate appropriateness for characterising range of cognitive functioning. Explore extent to which individual's performance on domain specific items fit expected pattern.	N = 186 community **TBI**; 140 male, 46 female; mean age 34 (range 18-60); mean education years 12 (range 6-20); median injury onset 32 months (range 1-336 months).	*Rating Scale Analysis (Rasch):* Several analyses performed to create unidimensional measure. Deleting easy and misfitting items improved measure without increasing error. Calibration of domains could only reliably distinguish normal/impaired.
Doninger, N. A., Ehde, D. M., Bode, R. K., Knight, K., Bombardier, C. H., & Heinemann, A. W. (2006). Measurement properties of the neurobehavioral cognitive status examination (Cognistat) in traumatic brain injury rehabilitation. *Rehabilitation Psychology, 51* (4), 281-288.	Calibrate item responses applying various RSA measurement strategies to investigate **Cognistat's** ability to distinguish meaningful levels of cognitive impairment in inpatient/outpatient TBI. Investigate appropriateness of **Cognistat** for characterising types of cognitive impairment.	N = 120 **TBI** inpatients drawn from consecutive admissions; 98 male, 22 female; mean age 37.5 (SD 12.6); 98 Caucasian, 11 African American; 22 educated > high school, 50 high school, 47 < high school; mean injury onset days 29 (SD 29). N = 286 community **TBI**; 230 male, 56 female; mean age 36 (SD 9.9); 69 Caucasian, 25 African American; 89 education > high school, 123 high school, 76 < high school; mean injury onset days 2,496 (SD 3,168).	*Rating Scale Analysis (Rasch):* Computed calibrations separately for inpatient/community groups using entire item, shortened item, and cognitive domains. Three strata of performance were differentiated despite a skewed distribution towards high performance among community sample. Elimination of easier items improved the instrument without increasing measurement error. Memory and verbal reasoning were the most difficult domains however analysis indicated significant memory. Calibration of domains could only reliably distinguish 2 levels (normal/impaired) in inpatient sample.
Duff, K., Beglinger, L. J., Jenks-Kettmann, J. D., &	Describes the performance of the **RBANS** and other	Case study 22 year old, left-handed, white, single woman with high school	All scores converted to standardised scores then standardised change scores calculated for comparisons. No

Appendix B Summary of included studies

Bayless, J. D. (2006). Pre- and post- right middle cerebral artery stroke in a young adult: A case study examining the sensitivity of the repeatable battery for the assessment of neuropsychological status (RBANS). *Applied Neuropsychology, 13 (3)*, 194-200.	neuropsychological tests: WAIS-III, RCFT, WRAT-3, TMT, TOMM, BDI-II.	education referred for cognitive assessment in context of history of numerous psychiatric admissions for self-harm, suicidality and mood lability. Diagnosed BPD and dysthymic disorder. Baseline IQ 85. 3 weeks after admission presented with left hemiparesis and dysarthria. MRI revealed right frontal embolic stroke.	significance testing. Post-stroke showed decline on Immediate memory index (story memory subtest), visuospatial/constructional index (figure copy & line orientation subtests), attention index (coding subtest, digit span improved). Most change in visuospatial/constructional index of RBANS. Performance on other neuropsychological tests consistent with RBANS.
Gaber, T. A-Z. K. (2008). Evaluation of the Addenbrooke's cognitive examination's validity in a brain injury rehabilitation setting. *Brain Injury, 22* (7-8), *589-593.*	Evaluate the practicalities, pattern of impairments, and establish the sensitivity of the **ACE-R** for TBI	N = 36 **TBI** admitted to rehab clinic; 31 male, 5 female; mean age 37.2 (SD 14.1); mean education years 11.1 (SD 2.7); mean injury onset months 22.5 (SD 12.5)	*Sensitivity* 100%/72%/56% (cut offs: 93, 88, 82). (Compared to MMSE 53%/36%/11%) *T-test:* Compared to norms published elsewhere, ACE-R total and subtest scores significantly more impaired ($p<0.001$).
Larson, E., Kirschner, K., Bode, R., Heinemann, A., Goodman, R. (2005). Construct and predictive validity of the repeatable battery for the assessment of neuropsychological status	Assess 1) convergent/divergent validity and 2) predictive validity of **RBANS**. Reference tests: TMT; line cancellation; EXIT; BDAE; WAIS-R vocabulary; RPM; Benton faces;	Study 1: N = 158 **stroke** patients; 80 female, 78 male; mean age 64.27 (SD 14.45); right hemisphere 49%, left hemisphere 44%, bilateral 7%. Study 2: N = 36 **stroke** patients; 19 female, 17 male; mean age 63.21 (SD 16.19); right hemisphere 67%, left	*Study 1) Convergent/divergent validity: Pearson's Product moment Bivariate correlations ($p<0.001$):* Attention index: not sig correlated with any measures attention but was sig correlated with measures of language (WAIS-R, BDAE). Language index: convergent validity with WAIS-R/BDAE but

in the evaluation of stroke patients. *Journal of Clinical and Experimental Neuropsychology 27 (1),* 16-32.	RBMT. Outcome measures: FIM motor/cognitive; FAI; CHART	hemisphere 25%, bilateral 8%.	also sig correlates with EXIT/RBMT. Visuospatial index: convergent validity with RPM/Benton faces but also sig correlations with TMT/line cancellation/WAIS-R/RBMT. Memory: convergent validity with RBMT but also sig correlations with EXIT/BDAE/WAIS-R. Controlling for verbal intelligence (WAIS-R) does not affect attention index but strengthens convergent validity for other indices. *Study 2) predictive validity: Pearson's Product moment correlations:* FIM-cog sig correlations with all RBANS scores except Attention. FAI significant correlation with visuospatial index.
McKay, C., Casey, J. E., Wertheimer, J., & Fichtenberg, N. L. (2007). Reliability and validity of the RBANS in a traumatic brain injured sample. *Archives of Clinical Neuropsychology, 22,* 91-98.	Evaluate the internal reliability of the **RBANS** index scores and construct validity of the subtest scores. Reference tests: CVLT; BVRT; MAE; COWAT; WAIS-III digit span/coding	N = 57 consecutive moderate to severe **TBI** patients referred to neurorehab outpatient clinic; 35 male, 22 female; mean age 35.72 (SD 14.62); mean education years 12.58 (SD 1.61); 40 Caucasian, 17 African American; mean injury onset months 84.88 (SD 101.15).	*Internal consistency: Chronbach alphas:* Immediate memory: 0.75; Visuospatial: 0.76; Language 0.33; Attention: 0.16; Delayed memory: 0.77; Total: 0.84. *Convergent validity: Correlation coefficients:* Significant correlations with tests measuring similar constructs. Range r=0.381 (List recog & CVLT recog) to r=0.83 (Coding & WAIS-III coding).

Appendix B Summary of included studies

McKay, C., Wertheimer, J. C., Fichtenberg, N. L., & Casey, J. E. (2008). The repeatable battery for the assessment of neuropsychological status (RBANS): Clinical utility in a traumatic brain injury sample. *The Clinical Neuropsychologist, 22* 228-241.	Examine clinical utility, sensitivity and specificity of **RBANS** in TBI compared to another clinical group. Hypotheses: 1) RBANS index and total score would be able to differentiate TBI patients. 2) Attention index would show largest impairment in TBI group.	N = 51 **TBI** consecutive referrals to rehab hospital; 28 male, 23 female; mean age 38.5 (SD 14.4, range 20-70); mean education years 13.1 (SD 1.9); 35 White, 16 Minority ethnic; injury onset months 63.3 (SD 92.3, range 1 – 312). Mostly moderate-severe TBI. N = 34 control group mixture of clinical/non clinical without brain injury; 13 male, 21 female; mean age 44.9 (SD 14.5, range 19-71); mean education years 13.8 (SD 2.8).	*T-tests:* Significant differences (P<0.001) on all index and total scores between TBI and control group, such that TBI performed worse. *Sensitivity/specificity/LR* RBANS cut-off <85 (1 SD) (<78 (1.5SD)) Immediate memory: *0.63 (0.33); 0.85 (1.0); 4.2 (33)* Visuospatial: *0.43 (0.28); 0.94 (1.0); 7.2 (28)* Language: *0.53 (0.29); 0.97 (1.0); 17.7 (29)* Attention: *0.69 (0.55); 0.79 (0.94); 3.3 (9.2)* Delayed memory: *0.67(0.45); 0.85 (0.91); 4.5 (5.0)* Total: *0.82(0.51); 0.94(1.0); 13.7(51).* *PPV/NPV/OCC:* RBANS cut-off <85 (1 SD) (<78 (1.5SD)) Immediate memory: 86.5 (100); 60.4 (50); 71.8 (60) Visuospatial: 91.7 (100); 52.5 (47.9); 63.5 (56.5) Language: 96.4 (100); 57.9 (48.6); 70.6 (57.6) Attention: 83.3 (93.3); 62.8 (58.2); 72.9 (70.6) Delayed memory: 87.2(88.5); 63(52.5); 74.1(63.5) Total: 95.5 (100); 78 (56.7); 87.1 (70.6)

Appendix B Summary of included studies

Nabors, N. A, Millis, S, R., & Rosenthal, M. (1997). Use of the Neurobehavioural Cognitive Status Examination (Cognistat) in traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 12,* 79-84.	Assess clinical utility and concurrent validity of **Cognistat** with other neuropsychological tests: TMT; token test; CVLT; logical memory WMS-R; Block design; and WCST	N = 45 **TBI**; 35 male, 6 female; mean age 39.5 (SD15.7); 82% African American; Mean education 10.4 years (SD 3.0); 47% mild, 18% moderate, 29% severe; mean injury onset 34.7 days (SD 25.5, range 6-66); 11 left hemisphere, 10 right hemisphere, 8 bilateral.	*Pearson zero-order correlations:* Cognistat total score related to education but not to injury severity variables. Cognistat attention – TMT: r = -0.33, p<0.05; Cognistat comprehension – token test: r = 0.3, p<0.05; Cognistat memory – CVLT: r = 0.68, p<0.001 and WMS-R: r = 0.43, p<0.05; Cognistat construction – Block design: r = 0.54, p<0.005. No significant association cognistat reasoning – Wisconsin.
Pachet, A. K. (2007). Construct validity of the Repeatable Battery of Neuropsychological Status (RBANS) with acquired brain injury patients. *Clinical Neuropsychologist, 21,* 286-293.	Examine psychometric properties of **RBANS** in comparison to extensive neuropsychological battery. Reference tests: CVLT; logical memory WMS-III; RCFT; digit-span; SDMT; WCST; GPT; FTT; TMT; Block design.	N = 37 **ABI** patients long-term rehab facility; 29 male, 8 female; mean age 42.65 (SD13.01, range 19-65); mean education years 11.59 (SD 1.99); 16 closed head injury, 14 cerebrovascular accident, 3 meningitis, 4 anoxic; mean injury onset months 9.75 (range 3-26); 90% severe injury.	*Pearson correlations:* All subtests (bar figure copy/recall - RCFT) reached significance, r = 0.61 – 0.78. Most index scores significantly correlated with reference tests at p<0.005 level except SDMT/TMT with attention; RCFT with immediate/delayed memory or visuospatial; TMT with visuospatial. 3 unmatched tests (TMT; FTT & GPT) correlated with one or more of RBANS subtest. WCST & block design did not correlate with any RBANS subtest.
Temple, R. O., Zgaljardic, D. J., Abreu, B. C., Seale, G. S., Ostir, G. V., Ottenbacher, K. J. (2009). Ecological validity of the neuropsychological assessment battery	Evaluate ecological validity of **NAB-SM by** assessing the relationship between the **NAB-SM** and the FIM in TBI patients.	N = 70; 57 male, 13 female; mod-severe **TBI**; mean age 36.0 (SD 13.6); mean education years 12.1 (SD 2.2); 47 Caucasian, 14 Hispanic, 8 African American, 1 Asian; mean injury inset 1.7 years (SD 4.0).	*Hierarchical regression:* The NAB-SM was significantly associated with FIM total, motor and cognitive scales over and above effects of sex and age. NAB total accounted for 26% of variance in FIM total, 11% FIM-motor, and 53% FIM-cognitive.

screening module in post-acute brain injury rehabilitation. *Brain Injury, 23,* 45-50.			NAB Spatial independently associated with FIM total and motor. NAB language, memory and spatial independently associated with FIM-cognitive.
Wallace, J. J., Caroselli, J. S., Scheibel, R. S., & High, W. M. (2000). Predictive validity of the neurobehavioural cognitive status examination (NCSE) in a post-acute rehabilitation setting	Examine 1) association between overall **NCSE (Cognistat)** and neuropsychological test performance. 2) Convergent validity of **NCSE** with similar neuropsychological tests. 3) Concurrent predictive validity of **NCSE** with neuropsychological tests. Reference tests: WAIS-R; MAE; JOLO; CVLT; WMS-R.	N = 48 severe **TBI** admitted to rehab service; 31 male, 17 female; mean age 29.8 (SD 11.3); mean education years 12.4 (SD 1.4); mean injury onset days 113.3 (SD 82.6).	*Correlation coefficients:* No. impaired performance on NCSE correlated with no. impaired performance on neuropsych tests: *r=0.56, p<0.001* *Kappa statistics:* Agreement between NCSE and neuropsych tests for indicating impairment = 0.79, kappa statistic 0.45, sensitivity 0.92, specificity 0.22 Significant correlations found for all paired tests, ranging from r=0.32 (Attention-WAIS-R digit span) to r=0.66 (Construction-WAIS-R Block design); except similarities-WAIS-R similarities and judgement-WAIS-R Comprehension. *Kappa statistics:* Classification agreement between paired tests poor except construction-WAIS-R Block design.
Wilde, M. C. (2006).The validity of the repeatable battery of neuropsychological status in acute stroke.	Examine interrelationships between index score and factorial validity of the **RBANS.** Reference tests: COWAT, VFD, CIM, MMSE.	N = 120 **ischemic stroke** patients admitted to rehab hospital; 106 females, 104 males; mean age 61.91 (SD 13.97); mean education years 12.27 (SD 3.01); 104 Caucasians, 88 African Americans, 16 Hispanic, 2 Asian; mean injury onset days 8.90 (SD 10.58); 77 left hemisphere, 92 right hemisphere, 41 bilateral.	*Pearson correlation coefficients:* All index scores were significantly correlated with each other and the total index score (r's 0.25 – 0.85, p's <0.001). *Factorial Validity:* 2-factor solution accounted for 61% variance: language/verbal memory factor and visuospatial/visual memory factor.

Appendix B Summary of included studies

			External validity: Lang factor correlated significantly with COWAT (r (43) =0.65, p<0.001) and MMSE (r (203) =0.57, p<0.001). Visual factor correlated significantly with COWAT (r (43) =0.30, p<0.05); VFD (r (26) =0.60, p<0.001) and MMSE (r (203) =0.52, p<0.001). *T-tests:* Right hemisphere performed significantly better on language factor and worse on visual factor than left hemisphere.
Wilde, M. C. (2010). Lesion location and repeatable battery for the assessment of neuropsychological status performance in acute ischemic stroke. *The Clinical Neuropsychologist, 24,* 57-69.	Examine the relationship between lesion side and location on the **RBANS** index and subtest performance. Hypotheses: 1) left hemisphere stroke would perform significantly better on visuospatial/constructional and attention indexes. 2) Right hemisphere stroke patients better immediate/delayed memory and language. 3) Subcortical better than cortical on language and immediate/delayed memory. 4) Cortical better than subcortical on attention and	N = 164 **ischemic stroke** (part of larger sample 471 consecutive admissions to rehab); 81 female, 83 male; mean age 61.28 (SD 13.94); mean education years 12.57 (SD 2.85); mean injury onset days 8.82 (SD 9.06); 78 Caucasian, 73 African American, 10 Hispanic, 3 Asian; 63 left hemisphere, 76 right hemisphere, 25 bilateral; 93 subcortical lesions, 70 cortical lesions.	*MANOVA:* statistically significant main effects for side and location, no interaction. Right hemisphere better language, attention, and immediate/delayed memory. Left hemisphere better visuospatial/constructional. Subcortical significantly better than cortical on visuospatial/constructional. Right hemisphere better: list learning, story memory, list recognition, semantic fluency, digit span. Left hemisphere better figure copy and line orientation. Subcortical better figure copy, line orientation, coding, story recall and figure recall.

	visuospatial/constructional.		
Zgaljardic, D. J., & Temple, R. O. (2010). Reliability and validity of the Neuropsychological Assessment Battery-Screening Module (NAB-SM) in a sample of patients with moderate to severe acquired brain injury. *Applied Neuropsychology, 17,* 27-36.	Provide preliminary data on the internal consistency and construct validity of the **NAB-SM** in patients with moderate-sever ABI. Reference tests: Matched: digit-span, tokens test, BNT, logical memory, visual reproduction, visual form discrimination test, block design, COWAT. Unmatched: TMT, digit-symbol coding, RAVLT, WCST.	N = 42 patients **ABI;** 31 males, 11 females; mean age 41.8 (SD 15.1); mean education years 13.4 (SD 2.7); 26 White, 9 Hispanic, 6 Black, 1 Asian; 24 TBI, 18 CVA; mean injury onset year 0.77 (SD 1.1).	*Internal consistency: Chronbach's alpha:* Total score 0.60, attention 0.39, language 0.4, memory 0.42, spatial -0.14, executive -0.37. *Construct validity: Pearson product moment correlation coefficients:* 10/19 significant correlations ($p<0.05$) between NAB-SM raw index/total scores and Neuropsych tests. NAB-Total correlated with at least one neuropsych test from each cognitive domain. Some evidence of convergent validity for all indices except executive function which did not correlate with any executive neuropsych tests. NAB-SM subtests correlated with matched neuropsych tests except shape learning & visual reproduction. Unmatched tests also correlated with exception of NAB digits forward and NAB visual discrimination.
Zgaljardic, D. J., Yancy, S., Temple, R. O., Watford, M. F., & Miller, R. (2011). Ecological validity of the screening module and the	Assess ecological validity of **the NAB (DL & SM**) in patients with moderate to severe TBI Reference test: MPAI-4	N = 47 (32 men, 15 women); mean age – 31.7 (SD – 11.4); mean educational attainment = 12.7 (SD = 2.3); 26 Caucasian, 16 Hispanic, 3 African American, 2 Asian; Injury onset mean	*Linear regression analyses:* NAB-SM Total index score significantly associated with MPAI-4 Total score, $F(1, 45) = 5.3$, $p=.026$. NAB-SM Total index score significantly associated with Ability subtotal score, $F(1,45)=5.3$, $p=.027$ and Participation

daily living tests of the neuropsychological assessment battery using mayo-portland adaptability inventory-4 in post-acute brain injury rehabilitation. *Rehabilitation Psychology, 56,* 359-365.		= 16.6 months (SD = 26.6), premorbid IQ mean = 97.4 (SD = 8.0). Diagnosis **moderate-severe TBI** (GCS 3-12, LOC >30 min &/or +ve neuroimaging)	subtotal score, F(1, 45)=7.8, p=.008. NAB-SM Cognitive index scores were significantly associated MPAI-Total and Ability and Participation scores (F's>3.0, p's<.022). NAB-DL significantly associated with MPAI-4 Total, Ability and Participation scores (F's>2.8, p's<.022). No associations with adjustment score of mpai-4.

MMSE = Mini-Mental State Examination; TMT = Trail Making Test; NAB = Neuropsychological Assessment Battery (DL = daily living, SM = Screening module); (M)TBI = (Mild) traumatic brain injury; ABI = acquired brain injury; GCS = Glasgow Coma Scale; LOC = loss of consciousness; MPAI-4 = Mayo-Portland Adaptability Inventory-4; FIM = Functional Independence Measure; ABI = acquired brain injury; ACE-R = Addenbrooke's Cognitive Examination-Revised; WMS (III/R) = Wechsler Memory Scales (three/revised); BIT = behavioural inattention test; ROC = receiver operating curves; PPV = positive predictive value, NPV = negative predictive value; AUC = area under the curve; LR = likelihood ration; PSI = person separation index; RPM = Raven's Progressive Matrices; RCFT = Rey-Osterreith Complex Figure Test; COWAT = Controlled Word Association Test; CVLT = California verbal learning test; RBANS = repeatable battery for the assessment of neuropsychological status; SDMT = symbol digit modalities test; WCST = Wisconsin card sorting test; GPT = grooved pegboard test; FTT = finger tapping test; EXIT = executive interview; BDAE = Boston Diagnostic Aphasia Exam; RBMT = rivermead behavioural memory test; WAIS (III/R) = Wechsler adult intelligence scale (three/revised); NCSE/Cognistat = neurobehavioural cognitive status examination; WRAT-3 = wide range achievement test-3; TOMM = test of memory malingering; BDI-II = Beck Depression Inventory-2; BPD = borderline personality disorder; FAI = frenchay activity index; CHART = Craig Handicap and Assessment Reporting Technique; BVRT = Benton visual retention test; MAE = multilingual aphasia examination; JOLO = judgment of line orientation; VFD = visual form discrimination test; CIM = complex ideational material; MANOVA = multiple analysis of variance; BNT = Boston naming test; RAVLT = Rey auditory verbal learning test.

Appendix C: Summary of SPANS subtests and indices

Index	Subtest (Total score)
Orientation Index (ORI)	Orientation to Person (4)
	Orientation to Time (8)
	Orientation to Place (4)
	Orientation to Condition (2)
	Orientation to Prime Minister/President (2)
	Time Estimation (2)
Attention and Concentration Index (ACI)	Digit Span Forward (6)
	Digit Span Backward (6)
	Sustained and Divided Listening I (10)
	Sustained and Divided Listening II (10)
	Counting Backwards (6)
	Monetary Calculations (8)
Memory and Learning Index (MLI)	Object Recall (6)
	Figures Recall (11)
	List Learning (18)
	List Recall (6)
	List Recognition (12)
	Word-Symbol Paired Associates (14)
Language Index (LAI)	Repetition (6)
	Naming (12)
	Yes/No Questions (6)
	Following Directions (6)
	Reading (6)
	Writing Sentences (5)
	Similarities (12)
Visuo-motor Performance Index (VPI)	Object Recognition (3)

Appendix C: Summary of SPANS subtests and indices

	Spatial Decision (12)
	Unusual Views (4)
	Figures Copy (16)
	Letter-Number Coding (12)
	Figures Recognition (3)
	Facial Expressions (4)
	3 and 1 Concept Test (16)
Processing Speed Index (PSI)	Letter-Number Coding (12)
	Counting Backwards (6)
	Spatial Decision (12)
	Monetary Calculations (8)
	Sustained and Divided Listening II (10)
Conceptual Flexibility Index (CFI)	3 and 1 Concept Test (16)
	Similarities (12)

THIS HAS BEEN REMOVED FROM THE ELECTRONIC COPY

THIS HAS BEEN REMOVED FROM THE ELECTRONIC COPY

Appendix F: Consent form

CONSENT FORM

Participant identification number for this trial:

Title of study: Norming and psychometric analysis of a cognitive test

Name of Researcher: Gerald H. Burgess, Psy, D., Clinical Psychologist

Please initial boxes:

☐	I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
☐	I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason
☐	I understand that the relevant data collected during the study may be reviewed by the primary investigator (Dr Gerald Burgess) and members of the research team with permissions of the primary investigator where it is relevant to the research, but that no personally identifying information will be attached to this data. I give permission for the primary investigator and research team to have access to this data.
☐	I agree to take part in the above study

_____	_____	_____
Participant	Date	Signature
_____	_____	_____
Person taking consent	Date	Signature

Visit the website http://www.pansa.synthasite.com for information relating to the progress and findings of this study

Appendix G: Means (standard deviations) and ranges for SPANS subtest/indices according to participant group and form

Group/Form	ABI		LTNC		Controls	
Subtest/index (Highest score possible)	A	B	A	B	A	B
Orientation to Person (4)	3.37 (.88) 2-4	3.17 (1.12) 1-4	3.86 (.45) 2-4	3.57 (.79) 2-4	4.00 (.00) 4-4	3.97 (.26) 2-4
Orientation to Time (8)	6.10 (2.24) 1-8	5.08 (3.03) 0-8	7.32 (1.68) 1-8	7.00 (1.92) 3-8	7.98 (.13) 7-8	8.00 (.00) 8-8
Orientation to Place (4)	3.45 (.94) 0-4	3.25 (.87) 1-4	3.86 (.45) 2-4	3.43 (.79) 2-4	4.00 (.00) 4-4	4.00 (.00) 4-4
Orientation to Condition (2)	1.61 (.64) 0-2	.92 (.90) 0-2	1.68 (.72) 0-2	1.71 (.49) 1-2	2.00 (.00) 2-2	2.00 (.00) 2-2
Orientation to Prime Minister/President (2)	1.28 (.69) 0-2	.83 (.91) 0-2	1.63 (.54) 0.5-2	1.14 (.90) 0-2	1.94 (.18) 1-2	1.73 (.51) 0-2
Time Estimation (2)	1.47 (.77) 0-2	1.67 (.65) 0-2	1.46 (.74) 0-2	1.29 (.95) 0-2	1.97 (.25) 0-2	1.89 (.32) 1-2
Digit Span Forward (6)	4.10 (1.52) 1-6	3.67 (1.97) 1-6	4.43 (1.35) 1-6	4.57 (.79) 4-6	5.25 (1.03) 2-6	5.44 (.65) 4-6
Digit Span Backward (6)	4.06 (1.84) 0-6	3.58 (1.68) 0-6	4.93 (1.12) 2-6	4.43 (.98) 3-6	5.23 (.82) 3-6	5.43 (.83) 3-6

Appendix G: Means (standard deviations) and ranges for SPANS subtest/indices according to participant group and form

Sustained and Divided Listening I (10)	9.02 (2.19) 1-10	7.25 (3.57) 0-10	9.29 (1.21) 5-19	10.00 (.00) 10-10	9.75 (.87) 6-10	9.92 (.28) 9-10
Sustained and Divided Listening II (10)	7.33 (2.86) 0-10	4.50 (3.53) 0-10	8.07 (2.23) 3-10	8.14 (1.46) 6-10	9.49 (1.09) 6-10	9.82 (.47) 8-10
Counting Backwards (6)	4.43 (1.53) 0-6	2.92 (1.88) 0-6	4.61 (1.64) 0-6	4.71 (1.60) 2-6	5.20 (1.08) 2-6	5.34 (.89) 3-6
Monetary Calculations (8)	5.55 (2.25) 0-8	4.17 (3.07) 0-8	6.00 (2.24) 1-8	6.29 (2.29) 3-8	7.03 (1.34) 4-8	6.98 (1.46) 2-8
Object Recall (6)	3.63 (2.10) 0-6	1.92 (2.19) 0-6	4.96 (1.04) 3-6	4.14 (2.55) 0-6	5.30 (.67) 4-6	5.31 (.56) 4-6
Figures Recall (11)	6.00 (3.97) 0-11	2.92 (3.32) 0-8	8.54 (3.25) 0-11	6.00 (4.28) 0-11	8.74 (2.14) 0-11	9.30 (1.53) 6-11
List Learning (18)	12.71 (3.48) 2-18	9.75 (4.56) 4-18	14.50 (2.35) 8-18	13.43 (2.36) 11-17	15.72 (1.64) 12-18	16.08 (1.43) 12-18
List Recall (6)	2.96 (2.12) 0-6)	1.75 (2.14) 0-6	4.93 (1.41) 1-6	3.86 (2.73) 0-6	4.89 (1.24) 2-6	5.43 (.85) 3-6
List Recognition (12)	10.51 (1.82) 5-12	8.17 (3.54) 0-12	11.54 (1.00) 8-12	11.29 (.95) 10-12	11.77 (.69) 9-12	11.89 (.45) 10-12

Appendix G: Means (standard deviations) and ranges for SPANS subtest/indices according to participant group and form

Word-Symbol Paired Associates (14)	9.08 (4.41) 0-14	4.75 (4.90) 0-13	11.39 (2.95) 4-14	11.43 (4.32) 3-14	12.95 (2.39) 6-14	13.61 (.80) 11-14
Repetition (6)	4.55 (1.53) 1-6	3.50 (1.58) 1-6	5.04 (.84) 3-6	4.57 (1.13) 3-6	5.52 (.62) 4-6	5.33 (.63) 3-6
Naming (12)	11.00 (2.47) 0-12	7.58 (4.50) 0-12	11.71 (.60) 10-12	11.00 (2.24) 6-12	12.00 (.00) 12-12	11.97 (.18) 11-12
Yes/No Questions (6)	5.57 (.76) 3-6	4.58 (1.31) 3-6	5.71 (.54) 4-6	5.57 (.79) 4-6	5.97 (.18) 5-6	5.97 (.18) 5-6
Following Directions (6)	4.94 (1.36) 1-6	3.75 (1.49) 1-6	4.89 (1.17) 2-6	5.29 (1.11) 3-6	5.87 (.34) 5-6	5.70 (.50) 4-6
Reading (6)	5.24 (1.39) 0-6)	3.58 (2.43) 0-6	5.64 (1.19) 0-6	5.57 (.79) 4-6	5.90 (.44) 4-6	5.90 (.44) 4-6
Writing Sentences (5)	3.51 (1.34) 0-5	2.17 (1.75) 0-5	3.79 (1.03) 0-5	2.71 (1.98) 0-5	4.56 (.62) 3-5	4.30 (.69) 3-5
Similarities (12)	9.16 (3.17) 0-12	6.00 (4.29) 0-12	10.68 (1.79) 5-12	10.86 (1.07) 9-12	11.36 (1.08) 8-12	11.21 (1.00) 8-12
Object Recognition (3)	2.51 (.65) 1-3	2.08 (1.24) 0-3	2.86 (.59) 0-3	2.42 (1.13) 0-3	2.93 (.25) 2-3	2.80 (.40) 2-3

Appendix G: Means (standard deviations) and ranges for SPANS subtest/indices according to participant group and form

Spatial Decision (12)	8.35 (2.80) 0-12	5.17 (4.04) 0-12	9.36 (2.63) 3-12	9.14 (3.02) 4-12	10.89 (1.94) 6-12	10.80 (2.10) 4-12
Unusual Views (4)	3.43 (.89) 0-4)	2.67 (1.56) 0-4	3.68 (.91) 0-4	3.43 (.79) 2-4	3.64 (.52) 2-4	3.75 (.51) 2-4
Figures Copy (16)	13.04 (3.42) 0-16	10.25 (6.17) 0-16	14.32 (2.47) 8-16	13.29 (2.14) 9-16	14.11 (1.71) 7-16	14.13 (1.25) 12-16
Letter-Number Coding (12)	5.94 (3.26) 0-12	3.67 (3.63) 0-11	7.68 (3.41) 1-12	6.71 (3.35) 1-11	11.34 (1.34) 7-12	11.44 (1.10) 8-12
Figures Recognition (3)	2.04 (.98) 0-3	1.67 (1.07) 0-3	2.61 (.69) 1-3	2.14 (1.07) 0-3	2.69 (.53) 1-3	2.80 (.51) 1-3
Facial Expressions (4)	2.90 (1.21) 0-4	2.17 (1.34) 0-4	3.14 (1.04) 0-4)	2.86 (1.07) 1-4	3.66 (.66) 1-4	3.67 (.60) 2-4
3 and 1 Concept Test (16)	12.51 (3.09) 4-16	9.33 (6.10) 0-16	14.18 (3.04) 4-16	15.43 (.79) 14-16	15.84 (.52) 14-16	15.59 (.78) 13-16

Appendix G: Means (standard deviations) and ranges for SPANS subtest/indices according to participant group and form

ORI (22)	17.28 (4.71) 6-22	14.92 (6.10) 5-22	19.80 (3.62) 5.5-22	18.14 (4.14) 12-22	21.89 (.33) 20-22	21.58 (.77) 19-22
ACI (46)	34.39 (8.75) 5-46	26.08 (12.42) 6-45	37.32 (7.13) 22-46	38.12 (5.43) 30-43	41.95 (3.31) 32-46	42.93 (2.34) 34-46
MLI (67)	44.90 (15.24) 14-66	29.25 (15.98) 10-61	55.86 (8.86) 32-66	50.14 (15.70) 24-62	59.36 (5.60) 34-66	61.61 (4.06) 51-67
LAI (53)	43.98 (9.53) 9-53	31.17 (13.62) 11-51	47.46 (4.22) 38-52	45.57 (7.81) 29-52	51.18 (1.77) 47-53	50.38 (2.01) 42-53
VPI (70)	50.71 (13.23) 9-69	37.00 (21.97) 0-63)	57.82 (11.50) 20-70	55.43 (9.29) 40-64	65.20 (3.27) 55-70	65.00 (4.22) 54-70
PSI (48)	31.59 (9.40) 0-47	20.42 (21.97) 0-63)	35.71 (10.21) 13-48	35.00 (8.96) 17-44	43.95 (3.48) 36-48	44.39 (3.55) 35-48
CFI (28)	21.67 (5.69) 6-28	15.33 (8.77) 2-27	24.86 (4.06) 14-28	26.29 (1.70) 24-28	27.20 (1.28) 23-28	26.80 (1.53) 21-28

Appendix H: Submission guidelines for Journal of Clinical and Experimental Neuropsychology

Manuscript preparation

1. Journal-specific guidelines

- Papers are accepted only in English. American English spelling and punctuation is preferred. Please use double quotation marks, except where "a quotation is 'within' a quotation".
- There is no word limit for manuscripts submitted to this journal. Authors should include a word count with their manuscript.
- **Abstracts** of 100 words are required for all papers submitted.
- **Abbreviations** that are specific to a particular manuscript or to a very specific area of research should be avoided, and authors will be asked to spell out in full any such abbreviations throughout the text. Standard abbreviations such as RT for reaction time, SOA for stimulus onset asynchrony or other standard abbreviations that will be readily understood by readers of the journal are acceptable. Experimental conditions should be named in full, except in tables and figures.
- **Colour charges.** Authors should restrict their use of colour to situations where it is necessary on scientific, and not merely cosmetic, grounds. 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 pages are £250 per figure ($395 US Dollars; $385 Australian Dollars; 315 Euros). If you wish to have more than 4 colour figures, figures 5 and above will be charged at £50 per figure ($80 US Dollars; $75 Australian Dollars; 63 Euros). Waivers may apply for some papers – please consult the Production Editor regarding waivers. Depending on your location, these charges may be subject to Value Added Tax .

2. General guidelines

- The style and format of the typescripts should conform to the specifications given in the Publication Manual of the American Psychological Association (6th ed.).
- All parts of the manuscript should be double-spaced, with margins of at least one inch on all sides. Number manuscript pages consecutively throughout the paper.
- Authors must adhere to SI units . Units are not italicised.
- **Section headings** should be concise and should not contain numbering.
- Authors should indicate whether their paper is a regular (original) article, a brief article, a case study or a review. Authors should include a word count with their submission.

Appendix H: Submission guidelines for Journal of Clinical and Experimental Neuropsychology

- **Manuscripts** should be compiled in the following order: title page; abstract; keywords; main text; acknowledgments; appendices (as appropriate); references; table(s) with caption(s) (on individual pages); figure caption(s) (as a list).

- **Title page.** This should contain only:
  (1) the title of the paper, the name(s) and address(es) of the author(s);
  (2) a shortened version of the title suitable for the running head, not exceeding 40 character spaces;
  (3) the name, address, email address, telephone, and fax numbers of one author to whom correspondence and proofs should be sent;
  The affiliations of all named co-authors should be 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 article is accepted.

- Each paper 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 .

- **Tables** should be kept to the minimum. Each table should be typed double spaced on a separate page, giving the heading, e.g., "Table 2", in Arabic numerals, followed by the legend, followed by the table. Make sure that appropriate units are given. Instructions for placing the table should be given in parentheses in the text, e.g., "(Table 2 about here)".

- **Results** of statistical tests should be given in the following form:
  "... results showed an effect of group, $F(2, 21) = 13.74$, $MSE = 451.98$, $p < .001$, but there was no effect of repeated trials, $F(5, 105) = 1.44$, $MSE = 17.70$, and no interaction, $F(10, 105) = 1.34$, $MSE = 17.70$." Other tests should be reported in a similar manner to the above example of an F-ratio. For a fuller explanation of statistical presentation, see the *APA Publication Manual* (6th ed.).

- **Abbreviations** that are specific to a particular manuscript or to a very specific area of research should be avoided, and authors will be asked to spell out in full any such abbreviations throughout the text. Standard abbreviations such as RT for reaction time, SOA for stimulus onset asynchrony or other standard abbreviations that will be readily understood by readers of the journal are acceptable. Experimental conditions should be named in full, except in tables and figures.

- **Acknowledgements** should be gathered into a brief statement after the correspondence. All sources of financial sponsorship are to be acknowledged, including the names of private and public sector sponsors. This includes government grants, corporate funding, trade associations and contracts.

- **Footnotes** should be avoided unless absolutely necessary. Essential footnotes should be indicated by superscript figures in the text and collected on a separate page at the end of the manuscript.
- **Biographical notes** on contributors are not required for this journal.
- For all manuscripts non-discriminatory language is mandatory. Sexist or racist terms should not be used.
- When using a word which is or is asserted to be a proprietary term or trade mark, authors must use the symbol ® or TM.

3. Style guidelines

- Description of the Journal's reference style
- Guide to using mathematical symbols and equations

4. Figures

- It is in the author's interest to 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 paper 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 paper (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 paper, and numbered correspondingly.
- The filename for a graphic should be descriptive of the graphic, e.g. Figure1, Figure2a.