Lenehan, ME, Summers, MJ, Saunders, NL, Summers, JJ and Vickers, JC

Does the Cambridge Automated Neuropsychological Test Battery (CANTAB) Distinguish Between Cognitive Domains in Healthy Older Adults?

Does the Cambridge Automated Neuropsychological Test Battery (CANTAB) distinguish between cognitive domains in healthy older adults?

Megan Elizabeth LENEHAN [a], Mathew James SUMMERS [b,c *], Nichole Louise SAUNDERS [b], Jeffery J. SUMMERS [a,d] & James C. VICKERS [a,b]

[a] School of Medicine, University of Tasmania, Australia
[b] Wicking Dementia Research & Education Centre, School of Medicine, University of Tasmania, Hobart, Tasmania, Australia.
[c] School of Social Sciences, University the Sunshine Coast, Australia
[c] Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom.

Word Count:

- Abstract = 246 words
- Manuscript = 4309 words
- Tables = 4

* requests for reprints should be addressed to: Assoc. Prof. M.J. Summers, School of Social Sciences (ML32), University of the Sunshine Coast, Locked Bag 4, Maroochydore DC, Queensland, Australia, 4558; tel +61 7 5456 3758; facsimile +61 7 5459 5767; email: msummers@usc.edu.au

Abstract

The Cambridge Neuropsychological Test Automated Battery (CANTAB) is a semi-automated computer interface for assessing cognitive function. We examined whether CANTAB tests measured specific cognitive functions, using established neuropsychological tests as a reference point. A sample of 500 healthy older (M = 60.28, SD = 6.75 years of age) participants in the Tasmanian Healthy Brain Project completed battery of CANTAB subtests and standard paper-based neuropsychological tests. Confirmatory factor analysis identified four factors: processing speed, verbal ability, episodic memory and working memory. However, CANTAB tests did not consistently load onto the cognitive domain factors derived from traditional measures of the same function. These results indicate that five of the six CANTAB subtests examined did not load onto single cognitive functions. These CANTAB tests may lack the sensitivity to measure discrete cognitive functions in healthy populations or may measure other cognitive domains not included in the traditional neuropsychological battery.

KEYWORDS: Cambridge Neuropsychological Test Automated Battery; aging; dementia; neuropsychological; measurement validity; confirmatory factor analysis.

Introduction

Computerised assessment in both clinical and research settings offer a number of potential benefits over traditional face-to-face assessment. Automated tests reduce administration time and costs, enhance the ease of data collection through automated scoring and normative comparisons, offer millisecond precision with response times as well as recording response times for all components of a task, have standardised procedures and enable the assessment of multiple cognitive domains in one testing session (Schatz & Browndyke, 2002). Additionally, with improving internet speed and connectivity, computerised assessment may enable regional, rural and remote communities' access to specialist neuropsychological services centred in metropolitan areas.

Numerous computerised test batteries are available to assess cognitive function, including: Mindstreams, CAMCOG, CNS Vital, and the Cambridge Neuropsychological Test Automated Battery (CANTAB).The CANTAB system was initially developed in the 1980s to assess cognitive function in the elderly and dementing populations (Robbins, James, Owen, Sahakian, et al., 1994; Sahakian & Owen, 1992).The CANTAB has established a large normative data set and has been widely used in clinical research, with over 1300 peer-reviewed papers supporting its use (Cambridge Cognition Limited, 2015). The CANTAB is a semi-automated test battery which can be administered on a laptop PC and more recently has been modified for administration on a handheld tablet. The CANTAB was designed to assess cognitive function in the elderly and dementing populations (Robbins, James, Owen, Lange, et al., 1994). The current release of CANTAB Eclipse comprises 25 tests designed to assess components of cognitive function which fall into 7 broad groups of tests: visual memory, executive function, working memory and planning, attention, semantic/verbal memory, decision making and response control, social cognition, and screening/familiarisation

(Cambridge Cognition Ltd, 2012). Tests can be run individually or in a customisable battery to enable users to test a series of participants on a set sequence of tests (Cambridge Cognition Ltd, 2012). The CANTAB has been utilised extensively in the study of various populations, including those with schizophrenia (Levaux et al., 2007), Parkinson's disease (Foltynie, Brayne, Robbins, & Barker, 2004), Huntington's disease (Lawrence et al., 1998), frontal and temporal lobe excisions (Owen, Roberts, Polkey, Sahakian, & Robbins, 1991) and normal functioning adults (De Luca, Wood, Anderson, & Buchanan, 2003; Robbins et al., 1998)

The construct validity of CANTAB has been largely based upon the ability of test measures to discriminate between normal adults and various clinical populations, including mild cognitive impairment (Klekociuk, Summers, Vickers, & Summers, 2014; Saunders & Summers, 2010), Alzheimer's disease (Saunders & Summers, 2010), epilepsy (Torgersen, Johan, Hans, Bernt, & Arne, 2012), ADHD (Gau & Shang, 2010) and various central nervous system diseases (Roque, Teixeira, Zachi, & Ventura, 2011). Despite its widespread application and ability to discriminate between clinical and normal functioning groups, the association between established neuropsychological measures and CANTAB components remains under-researched.

Factor analytic studies by Robbins and colleagues (Robbins, James, Owen, Sahakian, et al., 1994) have been conducted on the CANTAB, finding four distinct factors corresponding to plausible cognitive functions. These were learning and memory, speed of response, executive processes, and visual perceptual ability (Robbins, James, Owen, Sahakian, et al., 1994). However, further research is required to compare CANTAB tests with domain measures based on long-standing, empirically sound neuropsychological tests in order to bolster the validity of this automated battery. An examination of the literature to date indicates that only one study has attempted to validate CANTAB tests against established neuropsychological tests. In a principal components analysis, Smith, Need, Cirulli, Chiba-Falek, and Attix (2013) found that CANTAB tests tended to load onto different components than standard neuropsychological tests, despite being purported to measure corresponding cognitive domains. In most cases, the grouping of CANTAB tests on components was unrelated to plausible cognitive domains. The authors acknowledge that the homogenous nature of the sample, including the relatively young age of the sample (M = 33.1 years, SD = 18.6) and high education level (M = 15.9, SD = 2.5) may explain the restricted range of scores produced on tests. Consequently, it is possible that there may have been insufficient performance variation to discriminate between functions.

Smith et al. (2013) also conducted correlational analysis between principal components derived from an established neuropsychological battery and CANTAB tests. This revealed weak to moderate associations between CANTAB subtests and traditional measures, however, these associations were less consistent after controlling for age and education (Smith et al., 2013). The results also showed that although CANTAB tests purported to measure memory were moderately associated with the memory factor, the same tests were also highly correlated with executive function (Smith et al., 2013). The authors proposed that while CANTAB tests might be a reasonable measure of “general” cognition they may lack the ability to measure distinct cognitive functions (Smith et al., 2013). However, weak to moderate correlations are to be expected when the initial validity estimates of neuropsychological tests are only moderate. This pattern of shared association across various cognitive domains is consistent with the complex nature of cognition and neuropsychological testing in general, with scores on one test frequently correlating with other tests from a variety of domains. For example, Stroop interference scores correlate modestly with tests that measure inhibitory processes, working memory, conceptual ability and speed of processing (Strauss, Sherman, & Spreen, 2006). Such interrelationships most likely reflect the conceptualisation of cognitive and neuropsychological functions. Functions such as working memory, attention, and executive function encompass a range of overlapping sub-processes and consequently vary in the degree of interrelationship. Therefore, measures of such constructs are likely to be interrelated and lack a high degree of discriminatory power.

The present study aims to examine whether six CANTAB tests measure the three specific constructs they are purported to assess: episodic memory, working memory, and executive function. These cognitive domains were assessed in a large sample of healthy older adults using an established set of standard neuropsychological tests alongside a subset of CANTAB tests. It is hypothesised that the underlying cognitive domain structure of the selected CANTAB tests will be consistent with the theoretical domains they are purported to measure (see Table 1).

[INSERT TABLE 1 HERE]

Method

Participants

The initial sample comprised 565 adults aged between 49 and 79 years at the time of recruitment into the Tasmanian Healthy Brain Project (THBP) (Summers et al., 2013). Participants were English as a first language speaking and predominantly of Anglo-Saxon ethnicity. Participants underwent annual comprehensive neuropsychological assessment utilising an array of standard paper and pencil based clinical tests as well as a selection of computerised assessment tasks (Summers et al., 2013). Participants who presented with a medical, neurological, or psychiatric disorder that could potentially influence neuropsychological test performance were precluded from entry into the THBP. Participants with moderately elevated anxiety or depression symptoms (as assessed on the HADS) were excluded from the data set. The resulting sample of 500 valid cases (146 male, 354 females) aged 49-79 years of age (mean 60.28, SD = 6.75) was used for analysis. The project was approved by the Human Research Ethics Committee (Tasmania) Network and further details of the study protocol have been previously published (Summers et al., 2013).

Materials

The materials for testing fall under three categories: screening tests; traditional neuropsychological test battery; and CANTAB test battery (see Table 1). Tests used in the THBP were selected on the basis of the having established reliability and validity for the assessment of discrete cognitive functions (for complete project protocol see Summers et al., 2013). The THBP battery included six CANTAB tests purported to assess episodic memory, working memory, and executive functions. The tests selected to form the traditional neuropsychological battery in the present analyses reflect the same three cognitive domains. Tests to represent a fourth domain, verbal ability, were also included in the traditional test battery for the purpose of establishing discriminant validity.

Screening Tests

Screening tests were administered to ensure the absence of dementia and clinically significant depression or anxiety. The Dementia Rating Scale, 2nd edition (DRS-2; Jurica, Leitten, & Mattis, 2001) is a 38 item instrument which provides an objective measure of dementia severity, as well as screening for individuals with possible dementia (Jurica et al., 2001). The DRS-2 has excellent utility and validity in diagnosing dementia (Jurica et al., 2001). Participants selected to take part in the present study displayed a DRS-2 AEMSS score ≥ 9; which is above the cut-off for clinical dementia and is consistent with intact general cognition.

The Hospital Anxiety and Depression Scale (Snaith, 2003) is a 14 item self-report scale designed to measure states of anxiety (HADSa) and depression (HADSd). The anxiety and depression subscales provide valid and reliable assessments of the severity of emotional state, as well as screening individuals for potential emotional disorder (Snaith, 2003). Participants were excluded from analysis if their HADS score was above the recommended cut-off for a moderate-severe emotional disturbance (HADS score ≥ 10).

Traditional Neuropsychological Battery

The 13 tests which form the traditional neuropsychological test battery cover four cognitive functions: episodic memory, working memory, executive function and verbal ability. These tests were selected on the basis of established reliability and validity in assessing the constructs they are designed to measure (Strauss et al., 2006) and because they correspond to the three cognitive functions the selected subtests of the CANTAB are purported to measure (Table 1).

Verbal episodic memory was assessed using the Rey Auditory Verbal Learning Test (RAVLT; Lezak, Howieson, Bigler, & Tranel, 2012) and the Logical Memory test (LM; Wechsler, 1997). The RAVLT is a verbal list learning and memory test in which 15 words are presented repeatedly across five successive trials. The number of words recalled after an interference trial (RAVLT A recall) was used as the outcome measure in the present study. The LM test uses verbal presentation of two brief narratives which are then recalled after a brief delay and then a 30 minute delay. The number of story units recalled after the 30 minute delay (LMII) was the outcome measure for the present study.

Visual episodic memory was assessed using the Rey Complex Figure Test (RCFT; Strauss et al., 2006). The RCFT is designed to evaluate visuospatial constructional ability and visual memory. Participants are required to firstly copy a complex geometric figure and then reproduce it from memory following a five minute delay. The RCFT was used to assess visuospatial memory, an outcome measured that is based on the number of figure units recalled.

Working memory was assessed using the Digit Span (DSP) and Letter-Number Sequencing (LNS) subtests of the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III; Wechsler, 1997). In the DSP subtest participants are presented verbally with number sequences and are required to repeat them in the same order (DSP forward), and on a subsequent trial in the reverse order (DSP backward). In the LNS subtest participants are verbally presented with a series of interleaved numbers and letters and are then required to repeat these with the numbers first in numerical order and then the letters in alphabetical order. These subtests assess how many pieces of verbal information an individual can attend to and manipulate prior to recall (Lezak et al., 2012). For both DSP and LNS the outcome measure was the total number of sequences correctly recalled.

Assessment of executive function and processing speed involved the Digit Symbol Coding (DSC) subtest of the WAIS-III (Wechsler, 1997), Trail Making Test trail A (TMT-A) and Trail Making Test trail B (TMT-B) and the colour-word incongruent trial of the 24-item Victoria version Stroop Colour-Word Test (Stroop-C; Strauss et al., 2006).These tests involve a variety of components including attention, maintaining a goal, suppression of automatic responses, processing speed and mental flexibility (Strauss et al., 2006). In DSC, participants are provided a series of numbers and must write down a corresponding symbol as quickly as possible. TMT-A involves connecting 25 encircled numbers which are spread across a page and participants are required to connect the numbers in order with a continuous line (Strauss et al., 2006). In TMT-B, participants must alternate between 13 numbers and 12 letters, connecting a number to a letter to a number and so on, with a continuous line (Strauss et al., 2006). In Stroop-C, participants are required to name ink colours from a stimulus card of colour names, where the colour of the ink is incongruent with the printed colour name (e.g., “green” is written in red ink) (Strauss et al., 2006). The DSC is scored on the number of items completed within 120 seconds, for the TMT and Stroop the outcome measure is time to successful task completion.

Verbal ability was assessed using the Controlled Oral Word Association Test (COWAT), and the Similarities, Vocabulary and Comprehension subtests of the WAIS-III (Wechsler, 1997). These tests assess the ability to produce fluent speech but also measure more “executive” aspects of verbal behaviour, such as the ability to think flexibly and organise output (Lezak et al., 2012). The COWAT requires individuals to name words (not proper nouns) that commence with a specific letter, with three trials each of 60 seconds duration. The outcome measure for this variable was the total number of unique words produced across the three trials. The Similarities subtest requires participant to explain how the concrete and abstract relationships between two items with increasing task difficulty on each subsequent pair. In Vocabulary, participants provide definitions of words of increasing complexity to show their understanding and ability to articulate word meaning. The Comprehension subtest assesses a participant's capacity to use language to express understanding of social conventions and rules.

CANTAB Test Battery

The Paired Associates Learning (PAL) test assesses visual episodic memory and learning (Cambridge Cognition Ltd, 2012) and requires participants to recall the spatial location of a predetermined number of unique patterns within a display matrix. PAL 'Total errors (adjusted)' was used as the outcome measure, which indicates the total number of errors across all assessed problems and stages and adjusts the total score for incomplete or failed trials (Cambridge Cognition Ltd, 2012).

The Spatial Span (SSP) test measures working memory capacity and is a computerised version of the Corsi Blocks task (Cambridge Cognition Ltd, 2012). The test requires participants to remember and recall a sequential series of coloured boxes in the correct order. SSP 'Span length', the longest sequence correctly recalled by the participant, was the outcome measure used for this test (Cambridge Cognition Ltd, 2012).

The Spatial Working Memory (SWM) test assesses the ability to retain and manipulate spatial information in working memory (Cambridge Cognition Ltd, 2012). In this task, participants use a process of elimination to find a blue token hidden inside each one of an array of boxes. SWM 'Between errors' reports the number of times a participant reopens a box in which the token has previously been found, indicating a failure to recall (Cambridge Cognition Ltd, 2012), and was the outcome measured used for this analysis.

The Rapid Visual Processing (RVP) test assesses visual sustained attention (Cambridge Cognition Ltd, 2012), which is a component of executive function. Participants are required to detect target number sequences amongst a presentation of pseudo-random numbers. The outcome measure used, RVP A' is a measure of signal detection threshold that incorporates the probability of a correct 'hit' and the probability of a false alarm into a ratio score that reflects the accuracy of the participant detecting target sequences (Cambridge Cognition Ltd, 2012).

The Match to Sample Visual Search (MTS) test requires participants to rapidly and accurately match patterns across various spatial locations with task performance thought to reflect executive function (Cambridge Cognition Ltd, 2012). Participants are required to find the match of a target pattern from a range of similar options. The outcome measure used in this analysis was MTS 'mean correct reaction time' which reports the time taken to make correct responses.

The 5 Choice Reaction Time (RTI) test assesses attention, measuring the speed of decision making and response time which are features of executive functioning. Participants touch the screen where a target yellow spot appears in one of five locations. The RTI ‘five-choice reaction time’ records the time taken to respond to a stimulus.

Data Screening

The data was screened for missing values. Seven participants were excluded from analysis due to a high number of missing values arising from technical difficulties with computerised testing equipment at the time of data collection. Of a total of 9538 data points, a total of 7 missing values were identified across 6 participants. Given the small proportion of missing values (0.07%), the 7 missing values were replaced with the mean score for the six neuropsychological measures they corresponded to (TMT-A, TMT_B, MTS, DSC, RCFT and Stroop-C).

Data Analysis

Analyses were conducted using SPSS v21. Confirmatory factor analyses were carried out to confirm the underlying constructs of the traditional neuropsychological battery and the CANTAB test battery. Factors were rotated using the direct oblimin method because there were theoretical grounds to expect that the underlying cognitive functions would be correlated and retained based on Eigen values greater than one. Given the large sample size, item factor loadings of ≥.3 could be considered statistically significant (Hair, Anderson, Tatham, & Black, 1998). However, only factor loadings of ≥.4 were considered to have practical interpretability in the present study (Hair et al., 1998).

Results

Sample

Background and demographic characteristics are presented in Table 2. Participants in the current study were healthy older adults aged 49-79 years, were free from dementia, were of average intelligence and were not clinically anxious or depressed. The majority of the sample were female (70%). A series of independent samples t-tests revealed that male participants were older ($t_{(456)} = 2.44$, p. <.05), reported higher levels of depression ($t_{(456)} = 3.38$, p. <.01), and had lower performance on the DRS-2 ($t_{(456)} = -4.44$, p. <.001) compared to females. Correlations were conducted to examine whether any demographic factors related to test performance. Cohen's (1988) cut off values were utilised with only correlations of a moderate (≥.5) or large (≥.8) magnitude considered meaningful given the large sample size. This revealed a positive, moderate correlation between the WTAR estimate of full scale IQ and Vocabulary score.

[INSERT TABLE 2 HERE]

Confirmatory Factor Analysis of traditional neuropsychological tests

Initially, the factorability of items was examined using a number of recognised criteria. Firstly, it was observed that 12 out of the 13 tests correlated at least .3 with at least one other test. Secondly, the Kaiser-Meyer-Olkin measure of sampling adequacy was .78, above the recommended value of .60 (Hair et al., 1998) and Bartlett's test of sphericity was significant ($\chi^2_{(78)} = 1599.09$, $p < .001$). The diagonals of the anti-image correlation matrix (measures of sampling adequacy) were all above the .5 recommended minimum (Field, 2009). Based on these indicators, factor analysis was considered to be suitable with all 13 traditional neuropsychological tests.

A confirmatory factor analysis was conducted on the 13 traditional neuropsychological tests to confirm their underlying factor structure. Four factors had eigenvalues greater than Kaiser's criterion of 1 and in combination explained 45.25% of the variance. Eleven of the 13 tests had primary loadings greater than .4. The factor loading matrix after rotation is presented in Table 3. The tests that cluster on the same factor suggest that factor 1 represents Processing Speed and explains 24.59% of variance. Factor 2 was labelled Verbal Ability and explained 9.46% of variance. The third factor represented Episodic Memory and accounted for a further 6.22% of variance. The final factor was labelled Working Memory and explained an additional 4.98% of the variance. The Rey Complex Figure test of visual episodic memory and the Controlled Oral Word Association test did not load on any factor.

[INSERT TABLE 3 HERE]

Confirmatory Factor Analysis of CANTAB subtests

It was observed that 4 out of the 6 tests correlated at least .3 with at least one other test. Kaiser-Meyer-Olkin (.77) and Bartlett's test of sphericity ($\chi^2_{(15)}$ = 367.16, $p < .001$) were acceptable, and measures of sampling adequacy were all above the .5 recommended minimum. A confirmatory factor analysis was conducted on the 6 CANTAB tests separately to confirm their underlying factor structure. A three factor structure was specified a priori to reflect the three cognitive domains these tests were purported to measure.

The three factor model explained 32.23% of the variance in combination. Four of the 6 tests had primary loadings greater than .4. The factor loading matrix after rotation is presented in Table 1. The tests that cluster on the same factor suggest that factor 1 represents Working Memory and explains 25.98% of variance. Factor 2 was labelled Processing Speed and explained 4.36% of variance. The third component represented Episodic Memory and accounted for a further 1.89% of variance. The Match to Sample test and Rapid Visual Processing test did not load significantly onto any factor. Generally speaking, the expected factor structure was observed with tests loading onto factors reflecting the purported constructs. As expected, the Match to Sample was most associated with the Processing Speed factor though this test did not have a significant loading. The Rapid Visual Processing test was associated with the Episodic Memory factor which was an unexpected finding, but may reflect an underlying memory requirement of the task whereby participants are required to identify and respond to three different three digit sequences interspersed with pseudo-random numbers.

[INSERT TABLE 4 HERE]

Confirmatory Factor Analysis of both traditional and CANTAB tests

A second confirmatory factor analysis was conducted to incorporate each test from the traditional battery and the CANTAB battery to examine whether the tests from each battery that were purported to measure the same cognitive function corresponded under the same factor structure. For consistency with the outcome of the traditional neuropsychological confirmatory factor analysis, a four factor structure was specified a priori.

It was observed that 17 out of the 19 tests correlated at least .3 with at least one other test. Kaiser-Meyer-Olkin (.84) and Barlett's test of sphericity were acceptable ($\chi^2_{(171)} = 2421.48$, $p < .001$) and measures of sampling adequacy were within the recommended range. Based on these indicators, principal components analysis was considered to be suitable for all 19 tests in the neuropsychological and CANTAB batteries.

The forced four factor model explained 38.62% of the variance in combination. Thirteen of the 19 tests had primary loadings greater than .4. The factor loading matrix after rotation is presented in Table 4. The tests that cluster on the same factor suggest that factor 1 represents Processing Speed and explains 23.04% of variance. Factor 2 was labelled Verbal Ability and explained 7.11% of variance. The third component represented Episodic Memory and accounted for a further 4.78% of variance. The final component was labelled Working Memory and explained 3.69% of the variance. Three CANTAB tests (Match to Sample, Reaction Time and Rapid Visual Processing) loaded onto Factor 1 with statistical significance but not practical significance. However, the Controlled Oral Word Association Test, and the Spatial Working Memory and Spatial Span CANTAB tests did not load significantly onto any factor.

[INSERT TABLE 5 HERE]

Discussion

This study aimed to explore whether a battery of CANTAB subtests measured the cognitive domains they were purported to measure. The initial factor analysis aimed to confirm the underlying factor structure of a battery of traditional neuropsychological tests with well-established validity and reliability. The results largely confirmed the purported underlying factor structure of the selected traditional neuropsychological tests. The four factor structure produced, and the tests which loaded significantly onto each factor, reflect the cognitive domains these tests are theorised to measure. However, two tests (Rey Complex Figure test & Controlled Oral Word Association test) did not load significantly onto any factor. Performance on verbal fluency tests have been shown to involve executive processes, such as the ability to shift mental sets (Abwender, Swan, Bowerman, & Connolly, 2001). Similarly, recall score from the Rey Complex figure test has been shown to relate to planning ability (Bennett-Levy, 1984), which is another higher order function. Specific executive functions cannot be measured discretely by neuropsychological tests. Rather purported tests of executive function tap multiple components of executive function (e.g., planning, decision making, etc.) as well as one or more underlying cognitive processes (e.g. memory, language etc.). That different measures of executive function do not coalesce onto a single factor in a factor analysis may reflect the multifaceted nature of the construct “executive function” rather than a flaw in the tests developed to assess components of executive function.

The second factor analysis examined whether CANTAB tests purporting to measure domains that were similar to those measured by the traditional neuropsychological battery did in fact coalesce onto the same cognitive domains. The results somewhat contradict expectations. The Paired Associates Learning (PAL) test of visual episodic memory was the only test that loaded onto the same factor as the traditional tests purported to also measure of episodic memory. These findings support the specificity of the PAL to effectively distinguish episodic memory in healthy older adults. However, the remaining five tests did not load with practical significance onto any factor (Match to Sample, Reaction Time, Rapid Visual Processing, Spatial Working Memory & Spatial Span). There are two key explanations for why tests from the CANTAB battery did not fall under the four factor structure.

First, it is possible that these tests measure modalities of cognitive function that were not measured within the current design and as such, fall outside of the factor structure. For example, while Spatial Working Memory and Spatial Span are tests reported to measure working memory function (Cambridge Cognition Ltd, 2012), these tests did not load onto Factor 4, which appears to represent working memory. This suggests that the traditional neuropsychological tests and the CANTAB tests may be measuring different components of working memory. For instance, both the Digit Span test and the Letter Number Sequencing test are administered verbally, and as such may reflect a specific auditory-verbal working memory domain. The two CANTAB tests on the other hand, may reflect a purely visual working memory domain which may be functionally independent of auditory-verbal working memory.

A second possibility is that some tests have shared variance across a number of cognitive domains and as such measure multidimensional rather than specific cognitive function. Closer inspection of the factor loadings for RVP A’, reveals that this test loaded relatively equally across each of the four factors (though the loadings were not significant). This pattern suggests that the task requires multiple cognitive strategies for effective completion and as such probably assesses higher-order functions. The Spatial Span and the Spatial Working Memory tests from the CANTAB battery also display shared variance, loading onto multiple factors (though not significantly). Given these findings, it is also possible that certain subtests of the CANTAB battery are measuring a multidimensional cognitive process (such as executive processing), as opposed to specific cognitive domains.

Validation research examining CANTAB tests against established neuropsychological measures is limited. The present results are consistent with the findings of Smith et al. (2013) who report that CANTAB tests show an inconsistent pattern of association with domain-specific composites derived from a traditional neuropsychological battery. Whilst previous research indicates that the CANTAB is sensitive to age-related decline (Robbins et al., 1998; Robbins, James, Owen, Sahakian, et al., 1994), and to intellectual impairment in younger adults (Edgin et al., 2010), there is little evidence that the CANTAB is sensitive to inter-individual differences in healthy, well-educated, adults. Longitudinal studies indicate a curvilinear slope of age-related cognitive decline, suggesting more rapid decline of functions including episodic memory, processing speed and visuo-spatial skills after the age of approximately 70 years (Hedden & Gabrieli, 2004). Consequently, as the mean age of the sample in the current study is 60.28 years, it is possible that there was insufficient performance variation to discriminate between functions. A second analysis examining those participants aged 70 years of age or older was not attempted due to insufficient sample size (n = 50).

When considering the ability of CANTAB tests to measure the cognitive domains they are purported to measure, it is important to recognise that these tests were developed to differentiate healthy aging adults from dementing populations. Prior research confirms that CANTAB tests effectively discriminate between normal functioning individuals and various clinical populations including mild cognitive impairment (Saunders & Summers, 2010), Alzheimer's disease (Saunders & Summers, 2010), epilepsy (Torgersen et al., 2012), ADHD

(Gau & Shang, 2010), Down syndrome (Edgin et al., 2010), and various central nervous system diseases (Roque et al., 2011) and on the basis of clinical cut-off scores in those with epilepsy (Torgersen et al., 2012). However, as CANTAB tests were not specifically designed with the purpose of measuring discrete cognitive functions, it is not surprising that not all of the tests included in the current battery coalesce with the cognitive domains from a traditional battery.

The CANTAB tests examined in the present study were restricted to three cognitive domains. Future research should examine the all 25 subtests included in the comprehensive CANTAB battery, which are purported to assess six cognitive domains. While the present study examined CANTAB test performance against traditional neuropsychological test performance in older adults, it is important to note that this sample was well educated with an average of 14 years of education. Future studies may benefit from examining these associations in an older sample with fewer years of formal education. Furthermore, a broader range of traditional neuropsychological tests could also be considered.

CANTAB tests are designed to distinguish individuals with cognitive deficits from those with normal cognitive function and specify which area(s) of cognition these deficits occur. The findings from the present study confirm that while the Paired Associates Learning test appears to measure the cognitive domain it is purported to measure (episodic memory), the remaining CANTAB subtests are not specific enough to discern one cognitive function from another. It is important to emphasise that the present research suggests that CANTAB software may not be sensitive enough to discern distinct cognitive functions in healthy older adults. This finding does not discredit the clinical utility of CANTAB software. The program was designed for use in clinical populations, specifically dementia disorders in the older age group, and as discussed there are many studies which support the ability of CANTAB to discriminate between healthy and clinical populations. Future research is required incorporating a more comprehensive range of CANTAB tests to include all cognitive functions the battery is designed to assess as well as including a sample of adults from a broader range of levels of function.

Acknowledgements

The researchers thank the participants for supporting the research. Ms Lenehan received a University of Tasmania Postgraduate Research scholarship as well as a supplemental scholarship from the Wicking Dementia Research and Education Centre. This project is funded by a National Health and Medical Research Council (NHRMC) Project grant (1003645), as well as the JO and JR Wicking Trust (ANZ Trustees).

Disclosure of conflicts of interest

Dr M Summers reports personal fees from Eli Lily (Australia) Pty Ltd and grants from Novbotech Pty Ltd, outside the submitted work. All other authors report nothing to disclose.

References

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. Assessment, 8(3), 323-338. doi: 10.1177/107319110100800308

Bennett-Levy, J. (1984). Determinants of performance on the Rey-Osterrieth Complex Figure Test: An analysis, and a new technique for single-case assessment. British Journal of Clinical Psychology, 23, 109-119.

Cambridge Cognition Limited. (2015). Company, from http:///www.cambridgecognition.com/company

Cambridge Cognition Ltd. (2012). CANTABeclipse Test Administration Guide. Cambridge: Cambridge Cognition Limited.

De Luca, C. R., Wood, S. J., Anderson, V., & Buchanan, J.-A. (2003). Normative data from the CANTAB. I: Development of executive function over the lifespan. Journal of Clinical and Experimental Neuropsychology, 25(2), 242-254. doi: 10.1076/jcen.25.2.242.13639

Edgin, J. O., Mason, G. M., Allman, M. J., Capone, G. T., DeLeon, I., Maslan, C., . . . Nadel, L. (2010). Development and validation of the Arizona Cognitive Test Battery for Down syndrome. Journal of Neurodevelopmental Disorders, 2(3), 149-164.

Field, A. P. (2009). Discovering statistics using SPSS: And sex and drugs and rock 'n' roll. London: SAGE.

Foltynie, T., Brayne, C. E. G., Robbins, T. W., & Barker, R. A. (2004). The cognitive ability of an incident cohort of Parkinson's patients in the UK. The CamPaIGN study. Brain : A Journal of Neurology, 127(Pt 3), 550-560. doi: 10.1093/brain/awh067

Gau, S. S.-F., & Shang, C.-Y. (2010). Executive functions as endophenotypes in ADHD: Evidence from the Cambridge Neuropsychological Test Battery (CANTAB). Journal of Child Psychology and Psychiatry, 51(7), 838-849. doi: 10.1111/j.1469-7610.2010.02215.x

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multivariate data analysis (Vol. 5th). New Jersey: Pentrice-Hall.

Hedden, T., & Gabrieli, J. D. (2004). Insights into the ageing mind: a view from cognitive neuroscience. Nature Reviews Neuroscience, 5(2), 87-96. doi: 10.1038/nrn1323

Jurica, P. J., Leitten, C. L., & Mattis, S. (2001). Dementia Rating Scale-2 (DRS-2): Professional Manual. Florida: Psychological Assessment Resources Inc.

Klekociuk, S. Z., Summers, J. J., Vickers, J. C., & Summers, M. J. (2014). Reducing false positive diagnoses in mild cognitive impairment: The importance of comprehensive neuropsychological assessment. European Journal of Neurology, early view. doi: 10.1111/ene.12488

Lawrence, A. D., Hodges, J. R., Rosser, A. E., Kershaw, A., ffrench-Constant, C., Rubinsztein, D. C., . . . Sahakian, B. J. (1998). Evidence for specific cognitive deficits in preclinical Huntington's disease. Brain, 121 (Pt 7)(7), 1329-1341. doi: 10.1093/brain/121.7.1329

Levaux, M. N., Potvin, S., Sepehry, A. A., Sablier, J., Mendrek, A., & Stip, E. (2007). Computerized assessment of cognition in schizophrenia: Promises and pitfalls of CANTAB. European Psychiatry, 22(2), 104-115. doi: 10.1016/j.eurpsy.2006.11.004

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). Neuropsychological assessment (5th ed.). Oxford: Oxford University Press.

Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J., & Robbins, T. W. (1991). Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man.

Neuropsychologia, 29(10), 993-1006. doi: http://dx.doi.org/10.1016/0028-3932(91)90063-E
Robbins, T. W., James, M., Owen, A. M., Lange, K. W., Lees, A. J., Leigh, P. N., . . . Summers, B. A. (1994). Cognitive deficits in progressive supranuclear palsy, parkinsons-disease, and multiple system atrophy in tests sensitive to frontal-lobe dysfunction. Journal of Neurology Neurosurgery and Psychiatry, 57(1), 79-88.
Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., & Rabbitt, P. M. A. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. Journal of the International Neuropsychological Society, 4(5), 474-490. doi: 10.1017/s1355617798455073
Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB) - A Factor-analystic study of a large sample of normal elderly volunteers. Dementia, 5(5), 266-281.
Roque, D. T., Teixeira, R. A. A., Zachi, E. C., & Ventura, D. F. (2011). The use of the Cambridge Neuropsychological Test Automated Battery (CANTAB) in neuropsychological assessment: Application in Brazilian research with control children and adults with neurological disorders. Psychology & Neuroscience, 4(2+(Special+Issue)), 255-265.
Sahakian, B. J., & Owen, A. M. (1992). Computerized assessment in neuropsychiatry using CANTAB: Discussion paper. Journal of the Royal Society of Medicine, 85(7), 399-402.
Saunders, N. L. J., & Summers, M. J. (2010). Attention and working memory deficits in mild cognitive impairment. Journal of Clinical and Experimental Neuropsychology, 32(4), 350-357. doi: 10.1080/13803390903042379
Schatz, P., & Browndyke, J. (2002). Applications of computer-based neuropsychological assessment. Journal of Head Trauma Rehabilitation, 17(5), 395-410.
Smith, P. J., Need, A. C., Cirulli, E. T., Chiba-Falek, O., & Attix, D. K. (2013). A comparison of the Cambridge Automated Neuropsychological Test Battery (CANTAB) with "traditional" neuropsychological testing instruments. Journal of Clinical and Experimental Neuropsychology, 35(3), 319-328. doi: 10.1080/13803395.2013.771618
Snaith, R. P. (2003). The Hospital Anxiety And Depression Scale. Health and quality of life outcomes, 1(1), 29-29. doi: 10.1186/1477-7525-1-29
Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). A compendium of neuropsychological tests: Administration, norms, and commentary (3rd ed.). New York: Oxford University Press.
Summers, M. J., Saunders, N. L., Valenzuela, M. J., Summers, J. J., Ritchie, K., Robinson, A., & Vickers, J. C. (2013). The Tasmanian Healthy Brain Project (THBP): A prospective longitudinal examination of the effect of university level education in older adults in preventing age-related cognitive decline and reducing the risk of dementia. International Psychogeriatrics, 25(7), 1145-1155. doi: 10.1017/S1041610213000380
Torgersen, J., Johan, T., Hans, F., Bernt, A. E., & Arne, G. (2012). Clinical Validation of Cambridge Neuropsychological Test Automated Battery in a Norwegian Epilepsy Population. Journal of Behavioral and Brain Science, 2(1), 108-116.
Wechsler, D. (1997). Wechsler Memory Scale - third edition (WMS-III): Administration and scoring manual: The Psychological Corporation.

Table 1.

Test batteries and theoretical cognitive domains

Test	Abbrev.	Cognitive domain
Traditional battery		
Rey Auditory Verbal Learning Test	RAVLT	Episodic memory
Logical Memory II	LMII	Episodic memory
Rey Complex Figure Test	RCFT	Episodic memory
Digit Span	DSP	Working memory
Letter-Number Sequencing	LNS	Working memory
Digit Symbol Coding	DSC	Processing speed
Trail Making Test Part A	TMT-A	Processing speed
Trail Making Test Part B	TMT-B	Processing speed
Stroop Colour-Word Test	Stroop-C	Processing speed
Controlled Oral Word Association Test	COWAT	Verbal ability
Similarities	SIM	Verbal ability
Vocabulary	VOCAB	Verbal ability
Comprehension	COM	Verbal ability
CANTAB battery		
Paired Associates Learning	PAL	Episodic memory/learning
Spatial Span	SSP	Working memory
Spatial Working Memory	SWM	Working memory and strategy use
Rapid Visual Processing	RVP	Visual sustained attention (speed of response)
Match to Sample Visual Search test	MTS	Reaction time (Processing speed)
Reaction Time - 5 Choice	RTI	Speed of response (Processing speed)

Table 2. Background sample characteristics and neuropsychological test performance

	Whole sample	Sample by gender		
Variable	Mean (SD)	Male	Female	p. (t-test)
Background characteristics				
Age (years)	60.28 (6.75)	61.35	59.85	.150
Gender Female (%)	354 (70.5%)	-	-	
WTAR (est. FSIQ)	112.58 (5.08)	111.81	112.53	.157
DRS-2 AEMSS	12.17 (1.94)	11.36	12.21	< .001**
Education (years)	13.91 (2.71)	14.03	13.81	.357
HADS - Anxiety	4.95 (2.61)	5.24	5.36	.654
HADS - Depression	2.27 (2.01)	2.97	2.27	< .001**
Traditional test battery				
Similarities (raw)	26.66 (3.49)	26.47	26.39	.808
Digit Symbol Coding (raw)	71.92 (14.35)	67.55	72.75	< .001**
RAVLT A recall (raw)	11.28 (2.67)	10.32	11.64	< .001**
RCFT recall (raw)	22.29 (6.02)	23.26	21.45	.001**
LMII recall (raw)	30.35 (6.21)	28.06	30.96	< .001**
Digit Span total (raw)	18.65 (3.95)	18.82	18.46	.322
Letter-Number Sequencing total (raw)	11.63 (2.36)	11.53	11.59	.785
Vocabulary (raw)	56.77 (5.90)	56.22	56.40	.752
Comprehension (raw)	26.12 (3.32)	26.90	25.66	< .001**
COWAT total (raw)	48.81 (11.69)	47.47	48.93	.175
Stroop-C time (sec)	25.72 (6.96)	27.50	25.77	.014*
Trail Making test Part A time (sec)	27.70 (8.48)	29.16	27.71	.081
Trail Making test Part B time (sec)	59.28 (19.54)	62.61	59.17	.064
CANTAB test battery				
PAL total errors adjusted (raw)	17.12 (16.80)	20.95	16.32	.004**
SSP longest recall (raw)	5.75 (1.21)	5.77	5.71	.587
SWM between errors (raw)	25.84 (19.07)	24.37	27.18	.109
RVP A’	.91 (.05)	.91	.91	.826
MTS mean correct RT (msec)	2466.99 (773.14)	2503.83	2476.14	.708
5-choice RTI reaction time (msec)	350.14 (50.56)	342.47	354.55	.009**

Note: WTAR = Wechsler Test of Adult Reading; DRS-2 = Dementia Rating Scale-2; HADS = Hospital Anxiety and Depression Scale; RAVLT = Rey Auditory Verbal Learning Task; RCFT = Rey Complex Figure Test delayed recall; LMII = Logical Memory delayed recall; COWAT = Controlled Oral Word Association Test; Stroop-C = Stroop Interference Trial; TMT-A = Trail Making Test Part A; TMT-B = Trail Making Test Part B; PAL = Paired Associates Learning; SSP = Spatial Span; SWM = Spatial Working Memory; RVP = Rapid Visual Processing; MTS = Match to Sample; RTI = Reaction Time.

* = p. < .05; ** = p. < .01

Table 3.

Confirmatory Factor Analysis of Traditional Neuropsychological Tests

	Traditional battery factor structure			
	Processing Speed	Verbal Ability	Episodic Memory	Working Memory
TMT-A	**-.786**	.106	.074	.053
TMT-B	**-.641**	-.071	-.009	-.134
Digit Symbol Coding	**.629**	.012	.111	.002
Stroop-C	**-.413**	-.093	-.110	-.048
Vocabulary	-.018	**.750**	-.009	.123
Similarities	.085	**.688**	.023	-.137
Comprehension	-.110	**.655**	.012	.032
COWAT	.158	.256	.000	.194
RAVLT	-.044	-.126	**.829**	-.007
LMII Delayed Recall	.014	.111	**.603**	-.016
RCFT	.094	.051	.283	.057
Digit Span	-.036	.001	-.010	**.793**
Letter Number Sequencing	.062	-.032	.044	**.721**

Note: TMT-A = Trail Making Test Part A; TMT-B = Trail Making Test Part B; Stroop-C = Stroop Interference Trial; RAVLT = Rey Auditory Verbal Learning Task; RCFT = Rey Complex Figure Test delayed recall; COWAT = Controlled Oral Word Association Test. Values >.40 were considered to have loaded on the factor and are in bold.

Table 4. Confirmatory Factor Analysis of CANTAB Test

	CANTAB Battery Factor Structure		
	Working Memory	Processing Speed	Episodic Memory
Spatial Span	**-.683**	.029	.044
Spatial Working Memory	**.517**	.031	.079
Reaction Time	-.025	**.531**	-.038
Match to Sample	.109	.282	.157
Paired Associates Learning	-.003	-.035	**.627**
Rapid Visual Processing	-.171	-.185	-.326

Values >.40 were considered to have loaded on the factor and are in bold.

Table 5.

Confirmatory Factor Analysis of Traditional and CANTAB Neuropsychological Tests

	Factor Structure			
	Processing Speed	Verbal Ability	Episodic Memory	Working Memory
Digit Symbol Coding	**.683**	.005	.079	-.025
TMT-A	**-.648**	.163	-.054	.007
TMT-B	**-.527**	-.017	-.114	-.193
Stroop-C	**-.497**	-.105	-.064	.001
MTS	-.372	.024	-.044	-.067
RTI	-.349	-.041	.071	.033
RVP A’	.302	.147	.194	.207
Vocabulary	.006	**.770**	-.022	.110
Comprehension	-.120	**.637**	.047	.054
Similarities	.111	**.630**	.082	-.110
COWAT	.240	.281	-.068	.169
PAL 6 shapes adjusted	-.022	.097	**-.720**	-.009
RAVLT	-.009	.004	**.666**	-.063
LMII Delayed Recall	.005	.184	**.564**	-.043
RCFT	-.001	.032	**.459**	.076
SWM between errors	-.185	.013	-.251	-.213
Digit Span	-.078	.049	-.044	**.813**
Letter Number Sequencing	.055	.018	.020	**.691**
SSP	.193	-.052	.219	.254

Note: TMT-A = Trail Making Test Part A; TMT-B = Trail Making Test Part B; Stroop-C = Stroop Interference Trial; RAVLT = Rey Auditory Verbal Learning Task; RCFT = Rey Complex Figure Test delayed recall; COWAT = Controlled Oral Word Association Test; MTS = Match to Sample; RTI = 5 choice reaction time; PAL = Paired Associates Learning; SSP = Spatial Span; SWM = Spatial Working Memory; RVP = Rapid Visual Processing..

Values >.40 were considered to have loaded on the factors and are in bold.