The Cantab Cognition Handbook

1st Edition

A comprehensive guide to computerized cognitive assessment with Cantab, the world's leading cognitive assessment software for over 30 years.
About the authors

Cambridge Cognition is a leading global provider of computerized cognitive tests combining neuroscience and technology to optimize mental health through life.

Our cognitive testing software, Cantab, has become the world’s most validated and comprehensive cognitive assessment system, used to conduct sensitive language-independent touchscreen tests in pharmaceutical clinical trials and research projects for over 30 years.

Copyright

The copyright in all material published in this handbook by Cambridge Cognition, including all portions of the text, graphics and the selection and arrangement of the material is owned by Cambridge Cognition, unless otherwise indicated.

All rights reserved. The Material may not be reproduced or distributed, in whole or in part, without prior written permission of Cambridge Cognition. However, reproduction in part, by non-profit, research or educational institutions for their own use is permitted if proper credit is given, with full citation, and copyright is acknowledged. Any other reproduction or distribution, in whatever form and by whatever media, is expressly prohibited without the prior written consent of Cambridge Cognition. For further information, please contact:

info@camcoq.com
+44 (0)1223 810 700

US office
Cambridge Cognition
2750 Rasmussen Road
Park City
Utah
84098
United States

UK office
Cambridge Cognition
Tunbridge Court
Bottisham
Cambridge
CB25 9TU
United Kingdom
# Contents

**Introduction** ......................................................................................................................... 3  
An overview of Cantab ................................................................................................................. 3  

**Familiarisation** ......................................................................................................................... 6  
Motor Screening Test (MOT) ........................................................................................................... 6  

**Psychomotor speed** ................................................................................................................... 7  
Reaction Time (RTI) ......................................................................................................................... 8  

**Attention** .................................................................................................................................. 12  
Rapid Visual Information Processing (RVP) .................................................................................. 13  

**Episodic memory** ...................................................................................................................... 18  
Paired Associates Learning (PAL) ................................................................................................. 19  
Delayed match to sample (DMS) .................................................................................................. 26  

**Working memory and executive function** .................................................................................. 31  
Spatial Working Memory (SWM) ................................................................................................... 32  

**Cognitive testing: test recommendation by disease** ................................................................. 39  
**Cognitive testing: quality consideration** .................................................................................. 40  

**Study practicalities** .................................................................................................................... 42  

**Data analysis considerations** .................................................................................................... 44
Introduction

The Cantab Cognition Handbook offers a comprehensive introduction to computerized cognitive assessment with Cantab. It starts with an overview of Cantab software, and then is divided into a series of sections that focus on the key cognitive domains of psychomotor speed, attention, episodic memory, working memory and executive function.

Each section starts by providing an overview of the theoretical background to the cognitive domain and how it is affected in neurodegenerative, psychiatric and neurological disorders. Each section then describes and discusses the Cantab test or tests that have been used to assess the cognitive domain, with a succinct review of the existing literature relating to each Cantab test included. Evidence is also integrated from pharmacological studies, studies of patients with Alzheimer's disease (AD), mild cognitive impairment (MCI), depression, schizophrenia and attention deficit hyperactivity disorder (ADHD) as well as data from neuroimaging research to provide a convergent overview of the use and interpretation of the Cantab tests.

The text concludes with general information on study design, study practicalities and data analysis considerations, making this a useful resource at all stages of studies using Cantab. It is likely to be of interest not only to psychologists and neuroscientists, but also to anyone with an interest in using theoretically motivated and empirically validated cognitive tests in their research.

An overview of Cantab

The Cambridge Neuropsychological Test Automated Battery (Cantab) is composed of 25 tests and was created at the University of Cambridge in the 1980s. Cantab has been in continuous development for over two decades. Cantab has established sensitivity to a large number of drugs and disorders and has been extensively validated. To date, Cantab has been cited in over 1500 peer-reviewed articles (http://www.cantab.com/cantab/site/bibliography.acds).

Cantab tests were designed to be independent of language and culture, which makes them ideal for use in multi-national studies. The battery has been used in 50 countries in academic research and in multi-national clinical trials.

Due to the nature of the Cantab tests, pre-baseline participant training is not required, which reduces the time taken up by cognitive testing in your trial. There are instead built-in practice phases at the start of tests that allow participants to familiarize themselves with the tasks. A simple motor screening test is also used to familiarize participants with the system and to ensure that they are able to follow instructions and use the touch screen accurately.

Most Cantab tests are graded in difficulty and as a result have been used to test highly impaired patients as well as high-functioning individuals. Cantab tests have excellent test-retest reliability and parallel forms to help guard against learning effects over repeated testing sessions.

This guide introduces you to the key cognitive domains assessed by Cantab and the measures used to interpret the data. For each test, there is a summary of some of the key findings relating to imaging, pharmacological and clinical studies. At the end of the manual we provide an overview of general testing and data analysis considerations when using Cantab.

Cognitive domains

Sound understanding of the complexities of cognitive function is needed in order to overcome the challenge of reliable assessment. IQ tests provide one way of measuring intellectual functioning. However, despite being undoubtedly useful and important, IQ tests only provide a broad and static profile of what we mean by ‘cognition’. Furthermore, some intelligence tests, particularly those based on verbal knowledge, are relatively insensitive to impairments such as those seen in early Alzheimer’s disease.
Cognitive neuroscience now divides cognition into different neural systems controlling, for example, aspects of perception, attention, short-term memory, working memory, long term memory (e.g. facts and personal episodes), language and executive functions, such as planning, decision-making and impulse control. These different processes work together to gather, store and make sense of information from the environment, and use these representations to guide the production of thoughts and behaviour.

Any comprehensive battery of cognitive tests should provide an overview of these different stages of behaviour, starting at the most basic. For cognition to occur at all, there must be evidence of normal sensory or motor functioning; if the information does not reach the brain or cannot be expressed by it, then clearly cognitive functioning cannot be accurately assessed.

Four main domains of function are assessed in Cantab: psychomotor speed (page 6), attention (page 12), memory (page 18) and executive function (page 31). Together, these domains provide a comprehensive assessment of different levels of cognitive processing.

Each of these cognitive domains relies on a distinct but overlapping set of brain areas. They are therefore differentially affected by age, disease and pharmacological manipulation. Combined with brain imaging methodologies, many of the Cantab tests have also been shown to be useful for defining the specific network of brain structures responsible for their performance (e.g. the prefrontal cortex, parietal lobe and striatum in the case of visuospatial planning). Correspondingly, a failure on a task may suggest some form of impairment in that specific brain system.

Cognition in central nervous system (CNS) disease

For each section we provide a brief overview of Cantab data in a range of neurodegenerative, psychiatric and neurological disorders, which provide an indication of the sensitivity of the tests to impairment in clinical populations and sensitivity to treatment effects. The core disorders on which we focus are:

- AD/MCI
- Depression
- Schizophrenia
- ADHD

**AD/MCI**

Alzheimer’s disease (AD) and mild cognitive impairment (MCI) are characterised by an insidious onset with progressive cognitive decline. The earliest cognitive impairment is episodic memory. As the disease progresses, cognitive deficits become more global and include impairments in attention, semantic memory and executive function (reasoning, planning and cognitive flexibility). Cognitive deficits reflect damage to the basal forebrain resulting in loss of acetylcholine throughout the cortex and especially the temporal cortex.

Cognitive decline is one of the clinical hallmarks of MCI/AD. Cognition is consequently a key outcome measure in any trials testing new products aimed at halting or even preventing the progression of AD pathology.


**Depression**

Major depressive disorder is characterised by depressed mood and/or loss of interest or pleasure in life activities, causing clinically significant impairment in social life, work, or other important areas of every day functioning.
The disorder is typically associated with deficits of varying severity across a range of cognitive functions including attention, working memory and executive function, particularly on tests that require effortful processing. A further cognitive hallmark of the disorder is a negative bias in the interpretation of (and memory for) emotionally-salient information. Some cognitive symptoms persist even in remitted individuals while others are present predominantly during acute phases and may increase in severity with progression of the illness.


**Schizophrenia**

Schizophrenia is characterised by widespread cognitive dysfunction (affecting attention, visual and verbal memory, executive function and social cognition) in addition to positive and negative symptoms. Cognitive Impairment Associated with Schizophrenia (CIAS) is an important determinant of functional outcome and is inadequately treated by antipsychotic medication, and thus represents a major target for novel therapeutics.


**ADHD**

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterised by symptoms of inattention, hyperactivity, and/or impulsivity that interfere with the everyday functioning of affected patients. While it predominantly affects children and adolescents, some patients continue to experience persistent symptoms at adulthood, which often leads to high unemployment rates and drug abuse if left untreated.

Understanding how ADHD and its treatment affect cognitive function has clear implications for the diagnosis and management of this condition.


**Cognitive testing in non-CNS disorders**

There is a growing recognition that cognition is a potentially important endpoint for researchers interested in a range of diseases and interventions which do not directly target the central nervous system, but which nevertheless may affect cognitive function.

Although the brain is quite well-protected from many blood-borne chemicals by the blood-brain barrier, many small molecules, including agents that may be therapeutic for illnesses affecting other organs, may be deleterious to brain function. Examples include environmental chemical drugs used to treat cancer, asthma, high cholesterol, cardiovascular problems, diabetes, and drugs that affect the immune system.

Virtually all effective drugs have some ‘adverse side-effects’. These can be the product of a drug’s non-specific mechanisms of action or may simply reflect effects at other receptors distal from the therapeutic target. Many of these side effects are not serious. Other effects, such as sedation, represent a clear cost but have to be set against the major benefits of medication. However, other actions, which may be insidious, can in the long-term be so disadvantageous that the treatment has to be discontinued. All of these issues have to be considered during the initial screening of new compounds or trials involving them, as well as research studies that investigate the repurposing of compounds for new indications.
Disturbances of normal homeostasis can also result in undesirable effects on cognition if the supply of nutrients and oxygen required for cognitive function is disrupted. Because of the impact of potential treatments for non-CNS disorders on cognition, there is increasing interest in assessing cognitive function in these areas. Desirable features of cognitive tests for the purpose of testing the safety of compounds include sensitivity to disease and pharmacological manipulation. This makes Cantab well suited to such studies, this guide.

**General references and introduction to Cantab tests**


**Familiarisation**

**Motor Screening Test (MOT)**

The MOT is used to perform familiarisation and screening, and provides a simple reaction time measure. The goal of the MOT is not primarily to assess cognition but to provide a general indication that psychomotor function is intact enough to proceed with more sophisticated cognitive testing.

On this test, a lack of drug or group effect should be interpreted as indicating that any effects identified on the rest of the Cantab battery are not confounded by general effects of the drug or disease state on motor control or low level perception.

In addition, it allows researchers to screen for any problems with vision, movement or language comprehension that may prevent the participant from responding meaningfully to the subsequent tasks. For this reason, it is recommended that it be administered first.

**Test Description**

The Motor Screening test (MOT) is a task designed to familiarise participants with the touch screen and computer. The procedure consists of a series of crosses shown in different locations on the screen. The participant is asked to touch each cross using the forefinger of the dominant hand.

The test takes approximately 1 minute to complete.
Psychomotor Speed

The ability to receive, process and respond to input from the environment in an efficient way is essential for effective functioning, particularly in situations such as driving when rapid decision making and response are critical for safety. The concept of psychomotor speed encompasses a component of both physical movement and mental processing speed. Slowing in either one of these aspects can result in an increase of reaction times. It has been suggested that slower speed and processing underlie decreases in cognitive functioning with age associated with neurodegenerative processes.

One of the key features of Parkinson’s disease is slowing: this encompasses both motor slowing and cognitive slowing. Other conditions, such as depression and traumatic brain injury (TBI), are also associated with reductions in psychomotor speed, as are some drug effects.

Importantly, interpreting performance on more complex tests of psychomotor speed relies on intact performance at this more basic level of processing.

Key references:
Reaction Time Test (RTI)

Several Cantab tests have an element of processing speed, requiring participants to respond as quickly as possible. RTI makes response speed the focus of the task. The structure of the task enables slowing in movement to be dissociated from slowing in cognitive processing. In addition, by increasing the number of items the participant has to monitor (from 1 to 5 circles), the cognitive load of the task can be increased, thereby placing heavier demands on attention.

Description of the test

RTI is a measure of simple and choice reaction time, movement time and vigilance during simple and 5-choice reaction time trials. This task also permits measurement of anticipatory/premature responding and perseverative responding.

The participant must hold down a button until a yellow spot appears on the screen, at which point they must touch the yellow spot as quickly as they can. The spot appears in a single location during the simple reaction time phase and in one of five locations in the 5-choice reaction time phase.

Core outcomes

<table>
<thead>
<tr>
<th>Key Outcome Measures</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Simple or 5-choice reaction time</td>
<td>The median duration between the onset of the stimulus and the time at which the subject released the button. This measure is calculated from correct, assessed trials in which the stimulus could appear in one location only / any one of five locations</td>
</tr>
<tr>
<td>Median Simple or 5-choice movement time</td>
<td>The median time taken to touch the stimulus after the button has been released. This measure is calculated from correct, assessed trials where stimuli could appear in one location only / any one of five locations</td>
</tr>
</tbody>
</table>
Brain systems

Reaction time and attention are dependent on a network of brain areas including the prefrontal and parietal lobes, and are also dependent on neurotransmitter release from mid-brain structures such as the nucleus accumbens and ventral tegmental area.

Pharmacological manipulations

The Cantab RTI task is a direct analogue of the rodent 5-choice serial reaction time test (5-CSRT), one of the most well-studied animal behaviour paradigms. In the rat, 5-CSRT shows sensitivity to discrete lesion sites in the prefrontal cortex, to cholinergic lesions in basal forebrain (e.g. McGaughy, Dalley, Morrison, Everitt, & Robbins, 2002), and to several classes of compound. In humans, the Cantab RTI test shows differential Pharmacological sensitivity. For example, clonidine, but not guanfacine, impairs five-choice reaction time performance in young healthy volunteers (Jäkälä et al, 1999).

Age effects

With increasing age there is a gradual but significant decrease in processing speed (Figure 1).

Sensitivity to impairment in clinical populations

Dementia and MCI

Alzheimer’s Disease (AD)

Cholinergic dysfunction is evident in Alzheimer’s disease. This led to the prediction that this task would demonstrate impairment in AD, and this has now been demonstrated. For instance, Swainson (2001) compared performance on Cantab tests in AD to that of a non-demented, clinical population (patients with depression) and a group of patients with ‘questionable dementia’ (QD). They found that both AD and QD subjects were impaired on the memory tests. However, only the AD patients showed impairment in both accuracy and latency in the RTI test. Decline on the mini-mental state exam (MMSE) was also found to correlate significantly with choice reaction time.

A more recent study has confirmed these findings. In a study which compared performance of AD patients to those with Mild Cognitive Impairment (MCI) and normal controls, Saunders and Summers (2010) found that the AD group was significantly slower on choice reaction time than the all the other groups. Significant slowing was also seen on the simple reaction time measure in amnestic MCI (a-MCI) and AD groups, with the AD group also being significantly slower than the a-MCI group.

A pharmacological study of the cholinesterase inhibitor tacrine in AD patients found that while there was no significant effect on tests of memory (e.g. PAL or DMTS), there was a significantly beneficial effect on accuracy and latency in RTI performance (Sahakian et al., 1993). This suggests that the deficits seen in AD are attributable to the alterations in cholinergic neurotransmission in this disease.

Mild Cognitive Impairment (MCI)

The data from studies which have examined MCI alongside AD and normal ageing (e.g. Swainson 2001; Saunders and Summers, 2010), reviewed briefly above, suggest that
there are no, or limited, differences between controls and MCI patients in reaction time measures. This has been confirmed by studies which have focused on subtypes of MCI (e.g. Klekociuk & Summers 2014a, 2014b).

However, at longitudinal follow-up of 10 and 20 months, a-MCI groups showed significant slowing in both simple and choice reaction time. This significant slowing was not observed in control or non-amnestic MCI participants (Saunders and Summers, 2011). It is unclear whether this increase in slowing with time represents part of a trajectory to dementia in this MCI group. However, such impairment seems to occur relatively late compared to deficits in episodic memory (e.g. PAL).

**Psychiatric and mood disorders**

**Depression**
Depressed individuals frequently report cognitive slowing as a symptom. Despite this, a recent meta-analysis of cognitive function in depression showed that there was no significant impairment in RTI in currently depressed patients (Rock et al., 2013), although some authors have reported an improvement in psychomotor speed during remission (Egerhazi et al., 2013).

**Schizophrenia**
Impairments in reaction time are frequently observed in schizophrenia. These tend to be marked and are present even in medication-naïve cases with first-episode schizophrenia (Fagerlund et al., 2004).

Deficits are still present, but are not as profound in early-onset schizophrenia (Fagerlund et al., 2006). Interestingly, a longitudinal study of early onset schizophrenia demonstrated significant impairment in both reaction times and movement times relative to controls. However, over five year follow-up there was no significant worsening in the severity of this impairment (Jepsen et al., 2010), suggesting that this is an aspect of cognition which is differently affected in early onset schizophrenia, and remains relatively stable in this group.

**Neurological and movement disorders**

**ADHD**
Attentional processing is a core deficit in ADHD. Consistent with this, impairments have been observed in children with ADHD. For instance, Gau and Huang (2013) compared both children with ADHD and their unaffected siblings to typically developing controls. They found that those with ADHD performed worse in the RTI task after controlling for sex, age, co-morbidity, parental educational levels and IQ.

**Traumatic Brain Injury (TBI)**
Processing speed and psychomotor slowing are constantly reported in studies of TBI. Parry et al (2004) reported that the mean performance of their TBI group was significantly poorer than that of their control group on all measures derived from the RTI, including the simple reaction time, simple movement time, 5-choice reaction time and the 5-choice movement time. Performance of both TBI and control participants was correlated with metabolite concentration on magnetic resonance spectroscopy, indicating a relationship between indices of reaction time and brain structural integrity.

Sterr at el. (2006) demonstrated that Cantab RTI was significantly impaired in patients with symptomatic TBI relative to non-symptomatic TBI patients and healthy controls, and that there was a significant correlation between RTI speed and accuracy and the severity of post-concussion symptoms, suggesting that the task is sensitive to clinically meaningful indices of TBI severity.

**Parkinson’s Disease (PD)**
A 2004 study of RTI in PD patients found slowing of movement in patients off medication, but not on medication (Fern-Pollak 2004). However, an earlier study examining
RTI across a range of disease severity found that deficits in reaction time and accuracy increased as the severity of PD increased (Riekkinen et al., 1998). Furthermore, withdrawal of dopaminergic medication had a differential effect on reaction time, depending on the severity of the Parkinson’s disease, with those with the most severe PD showing the largest effect on reaction time. Choice reaction time performance has been found to be a significant predictor of freezing of gait, an important motor feature of PD (Shine et al., 2012).

References

- Gau, S. F., & Huang, W. L. (2014). Rapid visual information processing as a cognitive endophenotype of attention deficit hyperactivity disorder. Psychological medicine, 44(02), 435-446.


**Attention**

Attention can be seen as a gateway for information into the other cognitive domains, regulating the access of information from both external (perceptual) and internal (memories, knowledge) sources to later processing and storage stages. The analogy that is frequently used to describe attention is that of a spotlight – an area or object where processing resources are concentrated, sometimes at the expense of items outside the focus of attention. Attention is a limited resource which weakens as the number of items over which it is shared increases. Attention can also be characterised on the basis of its temporal characteristics. When asked to maintain attention over a long time, fatigue or boredom sets in and accuracy can be compromised. The capacity to resist distraction and maintain the focus of attention in the face of distracting information is referred to as “selective attention”, while the ability to maintain a steady level of attention over time is referred to as “sustained attention”.

Attention depends on a broad network of areas in the frontal and parietal lobes, including the dorsolateral prefrontal cortex and the anterior cingulate. Frontal regions are thought to be particularly critical in guiding the goal-directed allocation of attentional resources. In addition, subcortical structures such as the pulvinar and the superior colliculus are involved in the shifting of attention.

Consistent with this widespread anatomical network, multiple neurochemical systems are implicated. Noradrenaline, acetylcholine and dopamine all interact in complex ways to modulate the shifting, maintenance and control of attention.

Disruption to the spatial characteristics of attention can be most clearly seen in the example of unilateral neglect following a stroke that affects the parietal lobes. This results in a "bias" away from one half of space or one half of an object, such that the patient does not report items in the neglected side. Disruption to the temporal characteristics of attention is a hallmark of ADHD, where mind-wandering and poor classroom performance are attributed particularly to deficits in sustained attention.

**Key references**


Rapid Visual Information Processing Test (RVP)

While the RTI task provides an estimate of vigilance, RVP assesses more complex aspects of attention. In addition to the requirement to respond accurately and quickly to a target, participants have to discriminate what is and is not a target in a rapidly changing stimulus array. This heavily taxes sustained and selective attention as well as processing speed.

The A-Prime (A') outcome measure is based on signal detection theory, and provides a single index of how good a participant is at detecting a target against ongoing distraction.

Description of the test

RVP is a test of continuous performance and visual sustained attention. In a white box in the centre of the screen, single digits appear in a pseudo random order at a rate of 100 digits / minute. Participants must detect a series of target sequences (for example, 2,4,6; 4,6,8; 3,5,7) and register responses by touching button. Nine target sequences appear every 100 numbers.

Core outcomes

<table>
<thead>
<tr>
<th>Key Outcome Measures</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-Prime</td>
<td>A’ (A prime) is the signal detection measure of sensitivity to the target, regardless of response tendency (the expected range is 0.00 to 1.00; bad to good). In essence, this metric is a measure of how good the subject is at detecting target sequences</td>
</tr>
<tr>
<td>Median latency</td>
<td>The median response latency during assessment sequence blocks where the subject responded correctly</td>
</tr>
</tbody>
</table>

Cantab Rapid Visual Information Processing test

Figure 2: Key anatomical and neurotransmitter systems involved in attention and attentional control. Attention depends on at least three separate yet interacting neurotransmitter systems: noradrenaline (blue), acetylcholine (yellow) and dopamine (green). These modulate the activity of regions involved in attentional and cognitive controls, such as the prefrontal cortex (PFC), anterior cingulate (ACC), pre-supplementary motor area (pSMA) and posterior parietal (PPC) regions.
Brain systems

Performance on the RVP task is associated with activation in a network of brain structures, including the frontal and parietal lobes (Coull et al., 1995). RVP activates the classic fronto-parieto-occipital “attentional” network (Figure 3).

Pharmacology

RVP is sensitive to a range of sedating and stimulant effects (see Chamberlain et al., 2011 for a systematic review). For example, performance is worsened by diazepam (Coull et al., 1995) and improved by nicotine (Sahakian et al., 1989). Methylphenidate reduces response latencies on this task in healthy young volunteers (Elliott et al., 1997), but does not affect performance in healthy elderly males (Turner et al., 2003).

RVP is sensitive to monoamine manipulations, including clonidine (Coull et al., 1995). Dietary depletion of the dopamine precursor tyrosine impaired RVP performance in recovered depressed patients compared to patients treated with a balanced amino acid mixture (Roiser et al., 2005). While stimulants can increase the rate of commission errors on this task, methylphenidate has been found to improve performance in adults with ADHD (Turner et al., 2005). Nicotine improves performance in healthy volunteers and patients with Alzheimer’s disease (Jones et al., 1992).

Ageing

With age, there is a reduction in the speed and accuracy of processing.

Sensitivity to impairment in clinical populations

AD and MCI

Alzheimer’s disease

A number of different studies have demonstrated impairment in RVP in patients with Alzheimer’s disease. Swainson et al (2001) compared four groups of patients: controls, patients with AD, a group with ‘questionable dementia’ (QD) and a group of depressed patients. Both AD and QD subjects showed impairments in latency and accuracy on RVP. A more recent study demonstrated that RVP latency was significantly longer in AD compared to controls. However, there were no significant differences between MCI and AD (Saunders and Summers, 2010).

Pharmacological studies have demonstrated deficits in accuracy and latency in RVP at baseline in AD patients relative to controls. These deficits were ameliorated in a dose-dependent manner by the administration of nicotine (Sahakian et al., 1989; Jones et al., 1992).

Mild cognitive impairment

Individuals with MCI have been found to have significantly worse performance on the RVP task compared to controls (Klekociuk & Summers 2014c; Saunders and Summers 2011), with some differences between amnestic and non-amnestic MCI in terms of latency and accuracy indices (Klekociuk & Summers 2014b), and no significant difference between MCI and early AD (Saunders and Summers 2010).

In terms of differentiating MCI from normal ageing, RVP latency and accuracy were two of ten significant predictors which could be used to correctly identify participants with MCI from normal ageing (Saunders and Summers, 2011).
normal controls with a sensitivity of 80.0% and a specificity of 87.9% (Klekociuk et al., 2014C). However, longitudinal analysis has not demonstrated a steady decline in performance over a 10 and 20 month follow-up in patients with MCI. Indeed, there is some evidence for improved performance over time in some individuals (Klekociuk & Summers 2014a).

**Psychiatric and mood disorders**

**Depression**

Patients with bipolar disorder are impaired on RVP during both manic and remitted states (Clark et al., 2001, 2002). In major depressive disorder (MDD), some degree of sustained attention impairment is present during depressive phases, but these impairments appear to recover completely during euthymia (Clark et al., 2005).

A recent meta-analysis of studies of depressed participants which have used Cantab showed that there was a significant deficit in RVP relative to controls. This impairment was present even in unmedicated and remitted participants (Rock et al., 2013).

**Schizophrenia**

Deficits in patients with schizophrenia have been reported in a number of studies. Prouteau et al. (2004) explored the pattern of associations between visual cognitive performance and community functioning in patients with Schizophrenia. They found that RVP performance was associated with global function, adjustment to living and behavioural problems.

In addition to deficits in schizophrenia, reductions in performance on Cantab tests including RVP were observed during the prodromal phase of the disease (Bartók et al., 2005). The authors concluded that deficits in attention and cognition were present at the early stages of development of psychosis and that Cantab might be a useful tool for early detection in patients at risk of developing schizophrenia.

**Neurological and movement disorders**

**Traumatic brain injury**

Numerous studies have shown that sustained attention is impaired in patients with TBI (Chan, 2005; Malojcic et al., 2008; Parry et al., 2004). For example, Salmond et al. (2005) demonstrated that TBI patients exhibited significantly longer response times and reduced levels of target detection on Cantab RVP compared with healthy controls. The ecological validity of the task is supported by findings such as those of Sterr et al. (2006), who demonstrated that performance on Cantab RVP was significantly associated with symptom severity.

**Multiple Sclerosis**

Deficits in sustained attention are also observed in MS, where participants with relapsing-remitting MS (N=29) have slower response latencies on the RVP (Roque et al., 2011).

**ADHD**

Attentional deficits, particularly the ability to sustain attention, are a core deficit in ADHD. Consistent with this, sustained attention performance on RVP is significantly impaired in ADHD (Turner et al., 2005, 2004). Gau and Huang (2013) studied attentional performance in typically developing controls, children with ADHD and their unaffected siblings. They found that, compared with controls, both children with ADHD and their unaffected siblings had significantly higher total misses and a lower probability of hits in the RVP task. Deficits in sustained attention were associated with a longer duration of methylphenidate use and IQ, but psychiatric co-morbidity and current use of methylphenidate were not.
References

- Coull J.T., Middleton H.C., Robbins T.W., Sahakian B.J. (1995) Clonidine and diazepam have differential effects on tests of attention and learning, Psychopharmacology, 120, 322-332
- Gau, S. F., & Huang, W. L. (2013). Rapid visual information processing as a cognitive endophenotype of attention deficit hyperactivity disorder. Psychological medicine, 44(02), 435-446.
Episodic Memory

Episodic memory refers to the ability to store and retrieve information about specific events or “episodes”. The material stored about each event is a unique combination of items which occur at a specific time and place. An example is seeing somebody in the street and remembering that they are Mike and that you previously met him on New Year’s Eve at your sister’s house. This unique combination of item (Mike), place (your sister’s house) and time (a specific New Year’s Eve) comprise an episode. This type of memory is distinct from information remembered without such a context, for example our general knowledge or vocabulary, which is accrued throughout life without being “tagged” with a specific time and unique context. This latter form of memory is termed semantic memory.

Episodic memory is the basis for much of our personal autobiographical memory and is crucial for everyday life. Without the ability to store material from our present and retrieve information from the past, we become unable to function. The devastating effects of Alzheimer’s disease are in large part due to the erosion of this form of memory, as the disease progressively removes the capacity to form new memories. This leads to disorientation in time and place, and eventually to loss of even long-stored memories such as the identity of family and friends.

The neural basis of episodic memory has been studied since the famous case in the 1950s of Henry Molaison, or HM, a patient who underwent surgical resection of the medial temporal lobes and was rendered profoundly amnesic, unable to store new information beyond a few minutes, despite being still able to acquire new motor skills and having otherwise intact language, perceptual and cognitive skills. Since then, there has been a large amount of research in animals and patient populations, including imaging studies, confirming the key role that medial temporal lobe structures, particularly the hippocampus, play in the formation and retrieval of information. It is self-evident that these processes of episodic memory do not occur in isolation, but work in concert with and depend on input from perceptual areas and strategic coordination of resources from frontal regions involved in executive control.

Decline in episodic memory is an unfortunate characteristic of ageing. However, progressive and severe loss of episodic memory is not normal, and is considered a key feature of Alzheimer’s disease. This is due to the fact that Alzheimer’s pathology is first present in the hippocampus and medial temporal lobe areas, before inexorably progressing throughout the brain to finally affect all areas of the cortex (see Figure 4), at which point multiple, profound deficits are seen in a wide range of cognitive domains.

Figure 4: Smith et al., (2002) Progression of Alzheimer’s disease pathology. The earliest pathological changes are present in the medial temporal lobe, spreading at a later stage to other brain areas, particularly those connected with the MTL, resulting in more widespread cognitive changes.

Key references

Paired Associates Learning (PAL)

As described above, one of the hallmarks of episodic memory is the binding of separate pieces of information together, such as a person and a specific time or place. This makes tests which capture this feature particularly useful as tests of episodic memory. PAL achieves this by explicitly requiring the recall of a location which was previously paired with an object.

Description of the test

The task consists of a number of stages that the subject must complete in order. For each stage, boxes are displayed on the screen. These boxes open one at a time in a randomised order. One or more of them will contain a pattern. The patterns shown in the boxes are then displayed one at a time in the middle of the screen, and the subject must touch the box where the pattern was originally located.

If the subject makes an error on a stage, the patterns are re-presented to remind the participant of their locations. For each stage, the same patterns may be re-presented up to a set number of times. When the participant gets all the locations correct, they proceed to the next stage. If they cannot complete a stage correctly, the test terminates. At the easiest stages of PAL, there is a single pattern to remember the location of, and at the most difficult there are eight patterns.

The patterns are designed to be difficult to verbalise so that verbal rehearsal cannot be used as a strategy. Their nonverbal nature also allows the test to be used in international studies without the need for stimuli translation. Participants are allowed multiple attempts at each trial, with automatic re-presentation of all stimuli after each attempt until the stage is passed or terminated. The test is adaptive so that if a trial is not completed despite multiple attempts, the test automatically terminates and the error score calculated includes an adjustment for errors at those stages that were not attempted.

Administration time is approximately 8 minutes, but depends on the performance of the participant.

Core outcomes

Total errors (adjusted)

Definition: The number of times the subject chose the incorrect box for a stimulus on assessment problems but with an adjustment for the estimated number of errors they would have made on any problems, attempts & recalls they did not reach due to failing or aborting the test.
Cognitive domains assessed

Paired associates learning is designed to assess visual episodic memory and learning. Tests of convergent validity have shown that there is a modest correlation between the PAL and story recall (-.19 immediate; -.21 delayed recall. Smith et al., 2013). However, the verbal component to story recall does limit the comparability of these tasks. In terms of the ecological validity of the measures, there is strong correspondence between PAL and self-reported memory problems (Swainson et al., 2001), as well as other measures of memory.

Brain systems

Successful performance of the PAL test is dependent on functional integrity of the temporal lobe, particularly the entorhinal cortex, but is also affected by resections of the frontal lobe (Owen, Sahakian, Semple, Polkey, & Robbins, 1995). Impairment on the Cantab PAL test correlates with hippocampal volume loss in schizophrenia (Keri et al., 2012).

The PAL paradigm has been modified for use in the MRI scanner, allowing direct assessment of the brain areas involved in performance. This has consistently shown involvement of the medial temporal lobe, particularly the hippocampus and parahippocampal regions (de Rover et al., 2011; Owen et al., 1995; Figure 5). De Rover et al. (2011) further showed that there were differences in the pattern of hippocampal activation during the PAL in MCI patients compared to normal controls.

These data support the use of this measure as a sensitive marker of hippocampal function.

Pharmacological data

The PAL task was designed to be a close analogue to tasks used in both primate and rodent research (e.g. Bussey et al., 2012). In these models, performance has been found to be dependent on the function of the hippocampus which is fundamental to learning associations between objects and locations.

The PAL task has been used in a range of pharmacological studies, particularly those targeting cholinergic systems implicated in Alzheimer’s disease (e.g. scopolamine) in both animals (e.g. Taffe et al, 2002) and humans. A summary of the data from published pharmacological studies in healthy participants is presented below. These extensive data demonstrate the sensitivity of these measures to both pharmacological impairment and improvement in performance. Of particular interest is the presence of dose dependent effects of scopolamine in healthy controls.

Ageing

Consistent with the early development of the medial temporal lobes, episodic memory improves rapidly in early childhood and remains relatively stable through mid-life, before worsening in later life (Figure 6).

Gender and education have also been found to impact on error scores, such that women and those with higher educational attainment perform better on the task. However, there are no significant interactions between age, education and gender.

Figure 5: Significant bilateral hippocampal and parahippocamal activation during the encoding (A) and retrieval (B) of PAL in elderly subjects. Rover et al. (2011)
Sensitivity to impairment in clinical populations

AD and MCI

Alzheimer’s disease

The medial temporal lobes are susceptible to Alzheimer’s pathology, particularly early in the course of the disease (Smith et al., 2002) and problems with episodic memory can be the first warning signs of AD. The sensitivity of PAL to hippocampal function has led to the suggestion that it may be used for the early and differential diagnosis of AD (Blackwell et al., 2004; de Jager, Lesk, Zhi, Marsico, & Chandler, 2008). Across studies the sensitivity and specificity indices were 0.94 and 0.91 respectively for detecting AD compared to elderly non-demented controls. The predictive ability for detecting incipient AD in populations with mild cognitive impairment is discussed below.

As would be expected, the rate of decline in PAL performance in AD is greater than that seen on average in MCI or healthy elderly populations (Figure 7). For example, PAL total error score (adjusted) increased by an average of 24 errors over one year in patients with dementia of the Alzheimer’s type (Fowler et al., 2002).

Mild cognitive impairment

Differentiating MCI from normal ageing is challenging, since MCI represent a heterogeneous group, with only certain individuals progressing to Alzheimer’s or another underlying cognitive disease. Nonetheless, PAL scores have a sensitivity of 0.83 and a specificity of 0.82 in differentiating adults with MCI from healthy older adults (Chandler et al., 2008). In a study at the University of Tasmania, patients with amnestic MCI showed poorer performance on PAL than either controls or non-amnestic MCI patients at baseline. When reassessed 10 months later, patients with amnestic MCI had worsened by an average of four errors at the six-pattern stage. In contrast, PAL scores of patients with non-amnestic MCI, who are at lower risk for Alzheimer’s disease, changed by less than one error (Saunders & Summers, 2011).

An important issue in the study of MCI is predicting which participants will go on to develop AD and which will not. A meta-analysis of 19 longitudinal studies has reported that approximately 10% of MCI cases per year progress to develop dementia (Bruscoli & Lovestone, 2004). PAL has shown considerable promise in this regard. Swainson et al. (2001) examined the ability of PAL to distinguish AD from controls and those with questionable dementia (QD) over 24 months. At baseline, both AD and QD participants were impaired on PAL, and there was a significant correlation between MMSE and PAL. PAL six-pattern error score was able to classify group membership with 98% accuracy. At baseline, those with QD appeared to be split into two distinct groups: one whose PAL scores resembled those with AD and one that resembled control subjects. These patients were followed up 6-12 months later, by which time the QD subjects had split into two groups: those whose performance had declined similarly to AD, and those whose performance had remained stable. By 24 months, all patients with poor and deteriorating performance had received a diagnosis of AD.
Similar findings have been reported in a series of studies by Fowler et al. (1995, 1997, 2002). They demonstrate that PAL is able to predict which participants with memory problems will go on to have AD two years later. The authors found that while participants with poor performance and declining trajectories went on to have a diagnosis of probable AD, those with higher baseline scores and stable performance did not (Fowler et al., 2002. See Figure 7). Using PAL alongside a test of naming, patients with relatively high baseline scores remained dementia-free at 32 months, while all of those impaired on both tests at baseline went on to receive a diagnosis of probable AD (Blackwell et al., 2004). Using the PAL alongside the Addenbrooke’s Cognitive Evaluation (ACE), Mitchell et al. (2009) obtained a sensitivity to conversion to AD of 94%. This compared to other cognitive tests, such as measures of semantic memory and executive function, which showed very low sensitivity to conversion, ranging from 0-25%. These and similar findings have led to the suggestion that PAL may be a useful marker for the subsequent conversion of MCI to AD (Blackwell et al., 2004; Sahakian et al., 1988).

Recent studies examining the profile of impairment in different subtypes of MCI have found that participants with amnestic MCI (a-MCI) show poorer performance than those with non-amnestic MCI on PAL (Klekociuk & Summers 2014a; Saunders & Summers 2011). Together with other measures (including indices from RVP, SWM and MTS), PAL scores contributed significantly to the discrimination of MCI cases from controls (Klekociuk & Summers, 2014c).

Non-Alzheimer’s dementia

Less common subtypes of dementia (and other rarer neurodegenerative conditions) are associated with a range of cognitive, behavioural and functional deficits, and there are a number of different clinical diagnostic criteria, of which memory is a part. While memory assessment can be helpful in earlier stages of assessment in these diagnoses, it may not be a presenting symptom until later stages of other disease progression.

For Lewy Body Dementia (LBD), there is a study showing that patients performed poorly on the PAL task. Interestingly, in this study their performance was more severely impaired than patients with AD, which suggests that PAL may have sensitivity to memory symptoms associated with LBD (Galloway et al., 1992). Fronto-temporal dementia (FTD) is often a complex presentation and memory loss is not typically a key feature in the early stages. Nevertheless, from the few studies conducted, there is some evidence that PAL performance is impaired in the early stages of FTD presentation (e.g. Deakin et al., 2003; Lee et al., 2003).
Psychiatric and mood disorders

Depression

There is some debate regarding the sensitivity of PAL to depression. Swainson et al. (2001) showed that older adults with major unipolar depression had PAL scores similar to age-matched healthy controls, with little overlap between these groups and patients with mild Alzheimer’s disease. Similarly, Sweeney et al. (2000) did not find a significant difference in PAL performance between their young unipolar depressed patients and healthy controls, but did find that mixed/manic patients made more errors than healthy subjects. O’Brien et al. (1993) found that seasonal affective disorder (SAD) patients were significantly impaired in terms of number of trials to reach criterion on PAL, compared both to healthy controls and to their own performance once recovered.

A recent meta-analysis of studies using Cantab tests in depressed patients (Rock et al., 2013) brought together these and similar studies and revealed that, while PAL performance was reduced in medicated currently depressed patients, no significant change was seen in a smaller number of studies of those not currently medicated.

Schizophrenia

Patients with Schizophrenia show a level of impairment similar to that seen in AD (Gabrovksa-Johnson et al., 2003). In addition, there was a significant correlation between right hemisphere ventricle size and performance on the PAL in patients with Schizophrenia. Bartók et al. (2005) compared the performance of the patients to that of the Cantab standardization database. The performance of the prespsychotic patients was significantly lower compared to the healthy individuals on the PAL test (p<0.001), as well as in tests of frontal lobe function.

Neurological and movement disorders

Multiple sclerosis

MS patients performed significantly worse than controls on PAL. In addition, in MS patients performance was associated with levels of disease-specific markers (glutamate) in hippocampus, thalamus and cingulate (Muhlert et al., 2014). This again demonstrates the sensitivity of the test to pathology of the hippocampus.

Traumatic brain injury

TBI patients make more errors than controls on the PAL test (Salmond, Chatfield, Menon, Pickard, & Sahakian, 2005), and the degree of impairment is associated with increased atrophy in the hippocampal formation; this is measured by white matter density (Salmond, Menon, et al., 2006), again implicating this structure as a key area mediating successful performance. A later study examining changes in brain structure outside the hippocampus found a statistically significant negative correlation between measures of white matter integrity in the corpus callosum on diffusion weighted imaging and memory performance (r = -0.588, P = .005) (Holli et al., 2010).

References

• Bussey, T. J., Holmes, A., Lyon, L., Mar, A. C., McAllister, K. A. L., Nithianantharajah, J.,


Kéri, S., Szamosi, A., Benedek, G., & Kelemen, O. (2012). How does the hippocampal formation mediate memory for stimuli processed by the magnocellular and parvocellular visual pathways? Evidence from the comparison of schizophrenia and amnesic mild cognitive impairment (aMCI). Neuropsychologia, 50, 3193-3199


Muhler, N., Atzori, M., De Vita, E., Thomas,


Delayed match to sample (DMS)

Delayed matching to sample is based on paradigms developed for use in memory research with non-human primates. These were designed to tap memory processes in a non-verbal manner and control for any deficits in perception or attention which may be responsible for decreases in performance. This is achieved by having a simultaneous condition where the participant matches simultaneously presented stimuli to an identical item and an array of visually similar items. Increasing the delay between offset of the item to be remembered and the onset of the choices makes this task demanding of memory capacity, and therefore sensitive to deficits of the medial temporal lobe.

Description of the test

DMS is a test of simultaneous and delayed matching to sample. This test assesses visual matching ability and visual recognition memory.

Participants have to select from four similar complex visual patterns displayed at the bottom of the screen the one that is the same as a pattern displayed in the centre of the screen. In some trials, the choice patterns at the bottom of the screen and the sample pattern in the middle of the screen are displayed simultaneously whereas in others the choice patterns are only shown after the sample pattern (delays of 0, 4 or 12 seconds between the presentation of the sample pattern and the choice patterns).

Administration time: approximately 7 minutes

Core outcomes

<table>
<thead>
<tr>
<th>Key Window Measures</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent correct</td>
<td>The percentage of assessment trials during which the subject selected the correct box on their first box choice.</td>
</tr>
<tr>
<td>Median Correct latency</td>
<td>The median latency from the available choices being displayed to the subject choosing the correct choice on assessment trials where the subject’s first choice is correct.</td>
</tr>
<tr>
<td>Probability of error given error</td>
<td>This measure reports the probability of an error occurring when the previous trial was responded to incorrectly and is used in calculations of A’ and B” (Note that this is incalculable if &lt;1 errors made).</td>
</tr>
</tbody>
</table>
Cognitive domains assessed

A study looking at the convergent validity of the Cantab tests in a clinical sample confirmed that there was a significant association between delayed matching to sample and a pencil-and-paper measure of the ability to remember and match visual patterns (Torgersen et al., 2012). This measure was not correlated with indices of verbal memory or executive function.

Brain Systems

Many lesion studies in both humans and non-human primates have indicated that delayed matching to sample is primarily sensitive to damage in the medial temporal lobes (particularly hippocampus) and frontal lobes (Sahgal and Iversen, 1978).

Imaging work has confirmed the contribution of the medial temporal lobes and furthermore has demonstrated that this contribution is delay-dependent. Performance at longer delays is more dependent on the medial temporal lobes. At short delays, the task can be performed on the basis of perceptual features, and therefore is more dependent on visual processing areas, such as the occipital lobes (Elliot and Dolan, 1999).

These findings have been more recently confirmed in a study by Picchioni et al., (2007) which supports the greater involvement of medial temporal and frontal structures with increasing delay periods (Figure 8).

Pharmacological manipulations

Matching to sample performance shows established sensitivity to pharmacological manipulations. Dose-dependent reductions in performance were seen following administration of scopolamine. Reductions in performance were also seen following administration of the adrenergic agonist clonidine. Conversely, small but non-significant increases were seen following administration of the acetylcholinesterase inhibitor donepezil. Administration of modafinil also produced a modest increase in performance.

Ageing

The effects of ageing on DMS are illustrated in Figure 9 using published data in a large Chinese sample. These data show a gradual decline with age in performance on this task.

---

Figure 8: Picchioni et al., (2007). Increasing activation in the anterior cingulate (upper panel) and in the medial temporal lobe (lower panel) with increasing maintenance delay across conditions (12,000 ms > 4,000 ms > simultaneous). Results are superimposed onto a standard template brain.

Figure 9: Age effects on DMS from Lee et al., (2013). N=184
Sensitivity to impairment in clinical populations

**AD and MCI**

**Alzheimer’s disease**

As can be expected on the basis of the clear role the medial temporal lobes play in AD, performance on this memory task is affected early in the course of AD and correlates with Mini Mental State Examination (MMSE) performance (Swainson et al., 2001). The memory profile of AD patients shows a delay-dependent deficit. As the delay period increased, and correspondingly the medial temporal lobe demands increased, performance of the AD patients became increasingly poor (Abas et al., 1990).

Importantly, researchers have found that DMS can aid in differentiating AD from MCI (Saunders & Summers 2010) and depression in older adults (Abas et al., 1990). Interestingly, in the study group, although both groups were impaired in the delayed MTS condition, AD patients made more random distractor errors, and the depressed patients’ degree of response slowing was not related to task difficulty, unlike the control group and the AD group, where reaction times increased with task difficulty.

**Mild cognitive impairment**

Consistent with the memory impairments characteristic of MCI, impairments in MTS performance have been found in these patients compared to controls (Klekociuk & Summers 2014 c), although of less severity than those seen in AD (Swainson et al., 2001; Saunders & Summers 2010).

There are some differences in performance in different MCI subtypes, with non-amnestic MCI showing relatively intact performance (Klekociuk & Summers 2014 b).

**Psychiatric and mood disorders**

**Depression**

Depressed participants often report subjective cognitive impairment, and differentiating dementia and depression in older adults can be challenging.

A recent meta-analysis showed that impairment was observed in both unmedicated and medicated patients who were currently depressed (Rock et al., 2013). However, these deficits were found to resolve after remission of the depression.

Middle-aged adults with severe depression show impairments both in accuracy and latency on DMS (Elliott et al., 1996). Critically, DMS scores show a moderate but significant correlation ($r = 0.5$) with clinical measures of depression such as the Montgomery-Asberg Depression Rating Scale and the Clinical Interview for Depression (Elliott et al., 1996).

Depressed individuals with bipolar I disorder show normal performance in the simultaneous condition that assesses perceptual ability without memory demand, but show increasing impairment as the delay period increases (Rubinsztein et al., 2006). This effect has also been shown in bipolar patients during euthymic states (Rubinsztein et al., 2000).

An interesting feature of depressed individuals’ performance on this task is the significantly increased likelihood of making an error after an error, i.e. increased sensitivity to negative feedback. This effect is seen to a significantly greater extent in depressed individuals than in patients with other neurocognitive disorders, such as Parkinson’s disease and Schizophrenia, or neurosurgical patients (Figure 10: Elliott et al., 1996). This over-sensitivity to negative feedback is an example of a negative bias in information processing that is seen in depression.
With regards to the differentiation of AD and depression in older adults, Abas et al. (1990) found that elderly depressed patients were not impaired on simultaneous matching to sample but were as impaired as AD patients in accuracy and reaction time on the DMS task. Both the depressed and the AD patients also showed a delay-dependent deficit in memory. Their performance was different from that of AD patients in two respects, however. Firstly, the AD patients made significantly more random distractor errors than the depressed patients or the controls, with significantly fewer shape distractor errors than colour distractor errors (unlike the other groups, who made roughly equal numbers of each error type). Secondly, the depressed patients’ degree of response slowing was not related to task difficulty, whereas in the control group and the AD group, reaction times increased with task difficulty.

Neurological and movement disorders

Traumatic brain injury

Consistent with the impact that TBI can have on cognitive and attentional systems, Salmond et al. (2005) showed that TBI patients make significantly more errors and responded more slowly relative to healthy controls. However, Parry et al. (2004) found that scores were within normal limits. The differences between the two studies are probably accounted for by differences in the severity of impairment in this heterogeneous condition.

ADHD

There have been fewer studies examining performance in ADHD on this task. However, Gau and Huang (2014) found that, compared with controls, participants with ADHD performed worse in medial temporal lobe (MTS) tasks after controlling for sex, age, co-morbidity, parental educational levels and IQ. This may be secondary to the attentional problems evident in ADHD.

References

unipolar depression: the influence of perceived failure on subsequent performance. Psychological medicine, 26(05), 975-989.

- Gau, S. F., & Huang, W. L. (2014). Rapid visual information processing as a cognitive endophenotype of attention deficit hyperactivity disorder. Psychological medicine, 44(02), 435-446.


Working memory and executive function

Working memory can be defined as the ability to hold material in mind while that material is actively processed. This cognitive component is critical to many everyday activities such as mental arithmetic and following multi-step instructions and directions. A network of brain areas is involved in these kinds of tasks, consisting of parietal, temporal and prefrontal cortex. As a key cognitive skill dependent on a network of areas, it is impaired in a range of developmental and degenerative disorders, particularly those impacting on the prefrontal cortex and dopaminergic neurotransmission, such as schizophrenia and ADHD.

Executive function is perhaps the most complex and multifaceted cognitive domain. It is comprised of multiple sub-processes, only some of which will be addressed here. Broadly, executive function allocates attentional resources and coordinates the work of other cognitive systems to guide action to fulfil goals.

An example of this can be seen in shopping for food at the supermarket: the shopper has to formulate a goal, which is to buy what they need, employ their (episodic) memory of where things are to navigate through the supermarket, find what they need on crowded shelves of similar items (selective attention), keep track of what they have bought so far (working memory). They have to do all this while resisting the urge to buy items which would blow the budget, and behave in socially appropriately ways to their fellow shoppers and staff. At the end of all this, they have to find where they parked their car (episodic memory again).

The coordination of these individually complex tasks, planning, strategic thinking and inhibitory control are all key aspects of executive function. These components depend on the frontal lobes, with different sub-divisions playing a role in distinct aspects. Damage to the frontal lobes, following head injury, for example, is associated with impairments of executive function, which can have a devastating effect on the ability of individuals to perform in cognitively demanding or unstructured real-life situations, and to make decisions. Deficits are also seen in conditions such as schizophrenia which also affect the pre-frontal cortex.

Key references

Spatial Working Memory (SWM)

The Spatial Working Memory test taxes the key aspects of working memory: storage of information over short periods of time and the updating and manipulation of that information to guide action. It also provides a measurement of strategy use, thereby indexing executive function in addition to working memory capacity.

Description of the test

SWM measures the ability to retain spatial information and manipulate it in working memory. It is a self-ordered task that also assesses the use of strategy.

Participants must search for blue tokens by touching the coloured boxes to open them. The critical instruction is that the participant must not return to a box where a token has previously been found. The task becomes more difficult as the number of boxes increases.

Core outcomes

<table>
<thead>
<tr>
<th>Key Measures</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between errors</td>
<td>The total number of times the subject revisits a box in which a token has previously been found in the same problem (calculated for assessed problems only)</td>
</tr>
<tr>
<td>Strategy</td>
<td>For assessed problems with six boxes or more, the number of distinct boxes used by the subject to begin a new search for a token, within the same problem</td>
</tr>
</tbody>
</table>

Cantab Spatial Working Memory test
Brain systems
Data from patients with brain lesions demonstrates that SWM performance is impaired by damage to the prefrontal cortex, especially the dorsolateral prefrontal cortex (Manes et al., 2002; Owen, Downes, Sahakian, Polkey, & Robbins, 1990).

Versions of the task adapted for use in brain scanning techniques have been used to shed further light on the neural substrates mediating SWM task performance. Positron emission tomography (PET) indicated that the dorsolateral and mid-ventrolateral prefrontal cortices are particularly recruited during task performance (Figure 11: Mehta et al., 2000).

Figure 11. Control participants performing this task show activation in the VLPFC and DLPFC

Pharmacology
Numerous studies have suggested that SWM performance is sensitive to dopamine manipulation. For example, SWM is impaired by acute tyrosine depletion (Harmer, McTavish, Clark, Goodwin, & Cowen, 2001), and improved by D2 agonist administration (Mehta et al., 2001). Furthermore, the stimulant methylphenidate significantly improves SWM performance in healthy volunteers as well as various patient groups (Kempton et al., 1999; Mehta et al., 2000; Turner et al., 2005). By integrating imaging and pharmacological approaches, it was found that dorsolateral prefrontal cortex activation during SWM performance is dependent on dopamine modulation (Mehta et al., 2000).

This task has been used in a number of pharmacological studies of patients. For example, adults with ADHD show significantly improved performance on SWM following administration of methylphenidate but not modafinil (Kempton et al., 1999; Turner et al., 2005).

Ageing
Consistent with the protracted development of the prefrontal cortex thorough childhood and adolescence and its sensitivity to normal ageing, performance on the SWM task shows both clear developmental and ageing effects particularly in the between-search errors index (Figure 12).

Figure 12: normative trajectory of performance in the SWM task. Higher scores indicate worse performance

Sensitivity to impairment in clinical populations
Dementia and MCI
Alzheimer’s Disease
In addition to affecting episodic memory, AD affects working memory, which is dependent on frontal and temporal lobe structures. Patients have been found to perform worse than healthy controls on spatial working memory (Sahgal et al., 1992). The pattern of performance is such that there is a steep
decrease in performance as task difficulty increases in both AD and controls, although AD performance is worse than that of controls at all stages of the task (Figure 13). This highlights the suitability of this test even in cases with mild dementia, as performance does not reach floor levels early on in the task.

Figure 13: Sahgal et al., 1992

Mild cognitive impairment

SWM contributes to discrimination of MCI from normal ageing in combination with other measures of memory, attention and executive function (Klekociuk & Summers, Neurology, 2014).

Significant differences in working memory capacity have been found between subgroups of MCI, with those with a-MCI+ (amnestic MCI with additional deficits in other domains) making more errors than those with purely amnestic MCI (Klekociuk & Summers 2014b; Saunders & Summers 2011).

The strategy score also showed significant differences between MCI sub-groups, with again the a-MCI+ performing worse than the a-MCI group. (Klekociuk & Summers 2014a). SWM strategy scores were significantly higher at baseline compared to 10 month and 20 month follow-up periods (Saunders & Summers 2011).

Non-Alzheimer’s dementia

There is less data regarding the profile of cognitive performance in other forms of dementia. Nevertheless, there is some evidence to suggest that working memory is impaired.

Rahman et al. (1999) used Cantab to profile the cognitive dysfunction of patients with frontal variant fronto-temporal dementia (fvFTD). Small groups of fvFTD and age-matched controls (n=8 in each group) were compared on a number of Cantab tests: pattern and spatial recognition, spatial span, spatial working memory and set-shifting.

In this study, fvFTD subjects were relatively unimpaired on the spatial working memory task, although differences were found in more complex Cantab measures of executive function.

However, Sahgal et al (1995) assessed differences in cognitive profiles between subjects with AD and those with dementia of the Lewy Body type (DLB). They examined spatial working memory in a small subset of their sample (AD, n=8; DLB, n=8), and found a differential pattern of errors. Control subjects made the fewest between search errors, followed by AD patients, and DLB patients made the greatest. For within-search errors, DLB patients were significantly impaired in comparison with AD and controls (these two groups did not differ). This differential pattern of impairment on spatial working memory is thought to reflect dysfunctions in non-mnemonic processes mediated by fronto-striatal circuits, which are more severely damaged in DLB.

Psychiatric and mood disorders

Depression

A recent meta-analysis of studies of depressed participants which have used Cantab showed that there was a significant deficit in spatial working memory, relative to controls. This impairment was present even in unmedicated
and remitted participants (Rock et al., 2013), suggesting that this is potentially an important cognitive variable in depression.

**Schizophrenia**

Cognitive Impairment Associated with Schizophrenia (CIAS) is an important determinant of functional outcome, and working memory impairment is a core component of this.

Impairments have been seen across the range of chronicity and severity of schizophrenia.

In hospitalised patients with chronic schizophrenia, Pantelis et al. (1997) found deficits in SWM relative to healthy controls. In addition, comparison between patients with schizophrenia and those with frontal lobe lesions showed equivalent impairments on spatial working memory, with increased between-search errors. Patients with schizophrenia were unable to develop a systematic strategy to complete this task, relying instead on a limited visuospatial memory span.

Elliot et al. (1998) considered neuropsychological deficits in schizophrenic patients with preserved intellectual function (IQ >90) in order to examine the specificity of their cognitive impairment independent of global decline. On Spatial Working Memory, the patients showed impairment on both mnemonic and strategic components of the task, in contrast to temporal lobe patients (mnemonic impairment only) and frontal lobe patients (strategic impairment only) (Owen et al, 1995). Thus the data presented in this study confirm the many previous findings of mnemonic and executive dysfunction in schizophrenia, but do not suggest primacy of one over the other.

Deficits are present even earlier in the course of the disease, before the cumulative effects of medication have had an impact. Hutton et al. (1998) researched executive function in first-episode schizophrenia in 30 patients and 30 healthy volunteers matched for age and NART IQ. Patients with schizophrenia showed the greatest impairments on tests of executive functioning, including SWM.

The potential importance of SWM in the cognitive phenotype of schizophrenia is highlighted in a study of patients before the onset of diagnosed schizophrenia. Wood et al. (2003) explored the relationship between working memory and negative symptoms in patients who had been referred to the Personal Assessment and Crisis Evaluation Clinic, Melbourne, Australia, as being at high risk of developing psychosis. There were 38 high risk patients (9 of which later become psychotic at least 12 months from baseline assessment) and 49 healthy controls. The high risk group performed significantly more poorly than the healthy controls, and those who went on to develop psychosis performed more poorly than those who did not, but this did not reach significance. However, for the group that went on to develop psychosis, there was a significant association between their SWM errors score and their total score on the Scale for Assessment of Negative Symptoms.

**Neurological and movement disorders**

**Parkinson’s Disease**

SWM performance is impaired in Parkinson’s disease (Owen et al., 1997, 1992). This impairment is thought to be driven by dopaminergic dysfunction in PD. L-dopa withdrawal impairs SWM task performance in patients with Parkinson’s disease (
Figure 14: Lange et al., 1992). Neuroimaging evidence suggests that the beneficial effect of L-dopa is mediated by improved efficiency in the dorsolateral prefrontal cortex (Cools, Stefanova, R. Barker, Robbins, & Owen, 2002).

Multiple sclerosis

SWM performance has been found to be impaired in patients with multiple sclerosis relative to controls (Figure 15: Foong et al., 2000; Foong et al., 1997). Furthermore, performance at the more difficult 6-box and 8-box stages of this task correlated significantly with frontal lobe lesion load determined using MRI in patients with multiple sclerosis (Foong et al., 1997), again indicating the sensitivity of this measure to frontal lobe impairments.

Figure 15: Patients with multiple sclerosis show impaired spatial working memory relative to controls at all levels of difficulty (Foong et al., 1997).

Traumatic Brain Injury

Sterr et al. (2006) showed a trend for poorer Cantab SWM performance in patients with symptomatic TBI relative to non-symptomatic TBI patients and healthy controls. Performance on Cantab SWM was significantly associated with symptom severity. A study by McAllister et al. (2001) showed that TBI patients demonstrate different activation patterns of working memory circuitry on fMRI relative to healthy controls. A further review by McAllister (2006) suggests that patients with TBI have problems in the activation and allocation of working memory, and that these may be secondary to reduced catecholaminergic function.

ADHD

Both individual studies (e.g. Kempton et al., 1999; McLean et al., 2004), and meta-analysis (Chamberlain et al., 2011), have shown extensive impairments in spatial working memory in ADHD. These are evident both in terms of between-search errors and strategy scores.

Methylphenidate significantly improves performance on SWM in children and adults with ADHD as well as in healthy volunteers (Kempton et al., 1999; Mehta et al., 2004, 2000; Turner et al., 2005).
References

- McLean, A., Dowson, J., Toone, B., Young,


<table>
<thead>
<tr>
<th>Cognitive Domains Measured</th>
<th>Information Input</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practice</td>
<td>Attention</td>
</tr>
<tr>
<td></td>
<td>Sensorimotor function and comprehension</td>
<td>Processing and psychomotor speed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>General assessment of cognitive function</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluating cognitive effects in healthy volunteers</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alzheimer's Disease (AD)</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild Cognitive Impairment due to AD</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention Deficit Hyperactivity Disorder (ADHD)</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuromuscular Disease</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Sclerosis</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epilepsy</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke and Cerebrovascular disease</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parkinson's Disease</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traumatic Brain Injury</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huntington's Disease</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Down's Syndrome (adults)</td>
<td>● ● ●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obsessive Compulsive Disorder (OCD)</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* If repeat testing is required we recommend the Attention Switching Task (AST)
<table>
<thead>
<tr>
<th>Representation</th>
<th>Using information to guide behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Executive Function and Decision Making</td>
</tr>
<tr>
<td>Pattern Recognition Memory (PRM)</td>
<td>Spatial Working Memory (SWM)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cognitive testing: quality consideration

Characterising cognition depends on the reliability of the measures used. Reliability of cognitive testing is increased by reducing sources of error or bias in measurement. There are several potential sources of error which computerised cognitive testing address:

- Reducing errors of measurement. In standardised pencil and paper tests, reliability depends on highly skilled administrators to accurately record responses, measure reaction times and then code, record and score the responses. Errors at each stage of this process can impact reliability. Computerised assessment interfaces directly with the participant and automates these processes, reducing potential errors.

- Lengthier tasks can increase reliability by administering more trials. However, in populations such as children, and those with cognitive impairment, these additional trials increase testing time and place additional burdens on sustained attention. This may paradoxically reduce reliability when participants’ attention wanes or they tire. Computerised assessment provides short forms of tasks allowing appropriate testing time for the study population.

- The ability to implement tasks in which the condition or stimuli presented depends on participants’ previous responses is a further advantage, again reducing testing time, and maintaining participant engagement.

- In populations such as Parkinson’s disease or normal ageing, difficulties with motor control can also introduce measurement error and reduce participant’s ability to comply with testing. The use of intuitive and well-designed interfaces can facilitate this. Research has shown that computerized cognitive testing is acceptable to elderly participants, often being better tolerated than pencil-and-paper assessments (Fillitt et al., 2008; Collerton et al., 2008).

In addition to the reliability of measurement, another key requirement of cognitive tests is their sensitivity to impairment and to change over time. Computerised cognitive testing can optimise this in a number of ways:

- Floor or ceiling effects limit the ability to detect change or differences in scores, particularly in very able or impaired samples. Adaptive algorithms, which alter presentation of stimuli dependent on previous performance, can prevent this.

- The provision of parallel forms and the presentation of random stimuli to enable tasks to be repeated over time can increase sensitivity to longitudinal change by minimising practice effects (Semple & Link, 1991; Louis et al., 1999).

Extensive published data from intervention studies allows study-specific estimates of effect size to be generated, which will ensure studies are adequately powered to detect differences.

References


Study Practicalities

Does it matter if my subject is colour blind?

Most of the tests can be administered if the subject is colour blind. For example, PAL has unique patterns that are independent of colour, so scores on these tasks should not be affected. In the DMS, however, subjects need to be able to differentiate colours. However, the results would only be affected in severe cases of colour blindness, which are quite rare.

We would suggest doing a very quick screen for colour blindness with a subject on the first testing session using the Ishihara Colour Test. Alternatively, the practice stages on each task can be used to ensure the subjects can differentiate the colours used.

Can I use Cantab with people with limited cognitive functioning?

This depends on the level of impairment. The advantage of Cantab is that the majority of the tasks are language and culture independent, and the tasks have an intuitive game-like quality ideal for testing in populations with limited mental capabilities. The Cantab tasks have been used in individuals from specific disease populations that are characterised by very limited cognitive functioning, for example, Alzheimer’s disease. The normative database includes children as young as four and adults over the age of 90.

Can I use Cantab with people with physical disabilities?

This depends on the type and level of physical disability. To determine whether a subject can complete Cantab, we recommend the completion of the MOT task for screening purposes. This task simply involves the subject touching crosses that appear on the touch screen. From performance on this task, we can assess any problems with vision, motor control or comprehension that may result in the subject being unable to complete the remaining Cantab tests.

What if the subject refuses to finish the test?

The majority of Cantab tests are graded in difficulty and designed to terminate prematurely when a subject reaches their cognitive limit. By setting the subjects expectations and providing reassurance and encouragement, a subject should always be able to complete the tests.

Some examples of positive encouragement that can be used to motivate a subject to continue are as follows:

"Well done, this test is very hard, you are doing well."

"Not long to go; you have nearly finished this test."

If the subject refuses to complete the test, despite encouragement, then please use the abort function.

If a test is aborted, we would recommend you record a comment at the end of the testing session. The comment should explain the situation, mention that the subject refused to continue, any specific reasons why and which particular tests were not attempted or aborted. This is very useful when reviewing data.

Cambridge Cognition does not recommend that aborted data is included for analysis, and our normal internal procedure is to remove these test runs.

What if I need to rerun the test?

If you needed to abort a test during a testing session, e.g. because of a fire alarm, then you should rerun the test as soon as possible.
Can data from aborted tests be used?

We would recommend that data from aborted Cantab test runs is excluded from the final dataset and not include in the analysis.

The psychometric validity of data could be significantly altered by aborting a test. To provide consistency across sites and raters it is extremely important that the Cantab tests are run in full. The majority of the tests are graded in difficulty and designed to terminate prematurely when a subject reaches their cognitive limit, therefore there should be no need to abort manually. For example, in the PAL task, if the subject fails to recall all of the locations correctly after a set number of attempts the test will terminate. By following the script, setting the subjects expectations and providing reassurance and encouragement, it should not be necessary to abort a test.

If an individual test is aborted in a test battery, then all data from fully completed tests in the session can still be included in the analysis.

Can I allow breaks during Cantab sessions?

We would recommend you allow breaks during the Cantab sessions, depending on the length of the testing period, providing that the breaks are between tests and not during a test.

It is best practice to keep the breaks consistent so that they occur at the same point in the testing session for all subjects. However, this is not always practical and additional breaks may be given as and when required.

Can I allow food, drinks, and cigarettes during Cantab testing?

Cantab tests are very sensitive, and differences in Cantab task performance have been demonstrated from everyday substances, such as the amount of caffeine contained in one cup of tea. Cigarettes and certain foods have also been demonstrated to influence performance. Therefore we would advise that food, drinks and cigarettes are not allowed during the Cantab testing period. Water may be given if needed.
Data analysis considerations

The analysis of cognitive variables raises a variety of statistical issues, and the aim of this section is to address some of these issues and provide guidance when writing a statistical analysis plan prior to the study. Naturally, it is not expected that the recommendations made below will be appropriate in all cases, and it is strongly recommended that you seek specialist statistical advice.

However, it is hoped that by following the recommendations contained within this document, statistical analyses performed on Cantab data will be both readily interpretable and performed according to valid statistical principles.

Checklist

This checklist gives a number of important things that you should consider at the study design stage:

- Have you considered sample size and statistical power for the study?
- If you are planning to use composite measures in the study, have you discussed their construction with a statistician?
- Which factors do you plan to use as covariates?
- Are you planning one-tailed or two-tailed tests?
- How will you assess whether the assumptions of parametric analysis have been met and what are the alternative tests to be used should these assumptions not be met?
- Have you discussed the procedures for dealing with missing data with a statistician?

Power Calculations

The specific power calculation you need will depend on whether you have a between-subjects design (e.g. difference between two independent means) or within-subjects design (difference between two dependent means e.g. from matched pairs).

We recommend that you power using two-tailed testing with an alpha of 0.05 (probability of false alarm) and a power of 0.8.

To help, below is a table of power calculations, which is for between-subjects design. By Cohen’s definition, an effect size of 0.2 is considered small, an effect size of 0.5 is considered medium and an effect size of 0.8 is considered large.

Table 1: Power calculation table

<table>
<thead>
<tr>
<th>Tails</th>
<th>Power</th>
<th>Alpha</th>
<th>Effect size</th>
<th>N1</th>
<th>N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two</td>
<td>0.80</td>
<td>0.05</td>
<td>0.2</td>
<td>394</td>
<td>394</td>
</tr>
<tr>
<td>Two</td>
<td>0.80</td>
<td>0.05</td>
<td>0.5</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>Two</td>
<td>0.80</td>
<td>0.05</td>
<td>0.8</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Two</td>
<td>0.80</td>
<td>0.1</td>
<td>0.2</td>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td>Two</td>
<td>0.80</td>
<td>0.1</td>
<td>0.5</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Two</td>
<td>0.80</td>
<td>0.1</td>
<td>0.8</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>One</td>
<td>0.80</td>
<td>0.05</td>
<td>0.2</td>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td>One</td>
<td>0.80</td>
<td>0.05</td>
<td>0.5</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>One</td>
<td>0.80</td>
<td>0.05</td>
<td>0.8</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>One</td>
<td>0.80</td>
<td>0.1</td>
<td>0.2</td>
<td>226</td>
<td>226</td>
</tr>
<tr>
<td>One</td>
<td>0.80</td>
<td>0.1</td>
<td>0.5</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>One</td>
<td>0.80</td>
<td>0.1</td>
<td>0.8</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

Data preparation

Before commencement of statistical analysis, it is always recommended that the data are inspected and, if necessary, transformed such that appropriate statistical tests are employed for a given cognitive measure. Note that it is likely that different transformations will be necessary for different measures, and in general the procedures below ought to be performed on a variable-by-variable basis.

Non-normality, skew and kurtosis

As discussed below, it is usually preferable to employ parametric statistics for the analysis of cognitive measures, since they provide a number of advantages over non-parametric statistics. Analysing raw data (e.g., percent correct, reaction time) provides the most readily interpretable results from parametric
analysis. However, for many cognitive outcome variables the raw data will be skewed, and therefore not suitable for analysis with parametric tests, which, assuming a relatively low number of data points, only produce accurate results if data are normally distributed. This is particularly common for reaction time measures, where a few slow subjects may lengthen the right hand tail of the distribution (i.e. positive skew), or accuracy measures on easy stages of tests where most subjects perform very well but a few inaccurate scores may lengthen the left hand tail of the distribution (i.e. negative skew).

The preferred method to detect skew is to plot the residuals from the statistical analysis, for example using a box and-whisker plot, and the statistician should use their judgment and experience in this matter. However, since skewed raw data will generally result in skewed residuals, it may be possible to detect skew in the raw data. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. Data sets with high kurtosis (leptokurtic) tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis (platykurtic) tend to have a flat top near the mean rather than a sharp peak. A normal distribution is said to be mesokurtic.

If significant skew or kurtosis is detected, a transformation may render the data suitable for parametric analysis. However, it is vital to choose an appropriate transform for a given outcome measure, since inappropriately transforming data might either reduce sensitivity or increase the chance of spurious results.

The transformation is dependent on the nature of the skew, and detailed advice should be sought. As a general rule, for positively skewed data a log transformation is recommended, while a square transform may be able to correct for negative skew. However, it should be noted that transformation will change the nature of the treatment estimates produced from parametric analysis. For example, if a log transform has been employed, the treatment estimates will represent the ratio of one group’s score to that of another group, and not differences in the original cognitive measure.

Following transformation, the residuals from statistical analysis and/or raw data should be inspected to confirm that the skew has been removed successfully. In some cases it may not be possible to produce a normal distribution by transformation, and non-parametric analyses will be necessary (see below).

If using a regression to analyse data, it is the normality of the residuals (in relation to predictors) that is required, rather than normality of the raw scores. In these cases, and assuming an adequate sample size, it may be that removing extreme outliers (i.e. standardised residuals > +3 / -3) enables this assumption to be met.

Including covariates in an analysis

In general, the covariates to be included in the general linear model should be detailed in the analysis plan prior to statistical analysis. It is recommended that variables such as test site and cognitive battery order are entered as random-effects to conserve degrees of freedom and improve sensitivity, while demographic variables such as age, IQ and gender should be entered as fixed-effects. In a repeated measures design, if baseline scores are available for each cognitive outcome variable of interest, baseline score can also be entered as a fixed-effect.

Many studies, for example in pharmaceutical trials, employ longitudinal designs, where subjects are tested at repeated time-points following the commencement of drug treatment. In such cases, it is strongly recommended that a repeated measures mixed-model is used. This is the most sensitive method to determine whether a drug starts to have a beneficial effect at a particular time-point following treatment, and can also be helpful in dealing with missing data. Where appropriate, an autoregressive function should
be included to account for correlations between measures taken at successive time-points. It may also be useful to include a drug treatment x time interaction term in the model. However, the exact nature of the longitudinal model to be employed should be discussed with a statistician. In some cases, absolute differences between treatment groups may provide the most meaningful interpretation of the data. In other cases, change-from-baseline scores may be most informative. In the case where change-from-baseline score is preferred, it is still recommended to include baseline score as a covariate in the analysis to provide maximum sensitivity. This may seem counter-intuitive, since using a change-from-baseline measure effectively already includes the baseline term in the model. However, on many cognitive outcome measures, it is possible that both the change-from-baseline score and the post-treatment score may covary with the baseline score due to boundary effects. In such cases, including the baseline score will improve sensitivity by accounting for extra error variance in the analysis.

**Composite measures**

A number of approaches have been taken with respect to the generation of composite measures. In general, it is advisable to discuss the construction of composite measures prior to the commencement of the study. Composite measures can be useful, since they may provide greater sensitivity than individual test results by providing a more precise estimate of an underlying cognitive process. Furthermore, composite measures can reduce the number of primary end-points in a trial, thus at least partially circumventing the issue of multiple comparisons. However, in some cases, a particularly focused cognitive measure will prove a more sensitive index than composite measures.

If it is intended that composite measures are employed, it is vital that the method of calculating the composite measure is considered prior to data collection. Furthermore, it is strongly recommended that only those composite measures based on published factor analyses that have previously been validated should be employed. If composite measures are calculated without using reference to established methods, the results may be difficult to interpret, since a change in the composite measure cannot be compared to other published studies. This makes the magnitude of a given change in a composite measure considerably more difficult to interpret with respect to previous literature. Factor analyses of Cantab scores in large datasets have been published (e.g., Robbins et al., 1994; see also Harrison et al., 2007 for non-Cantab tests), and should be examined for guidance and to guide the analysis strategy.

**Test-retest reliability**

Technically, reliability measurement is an attempt to estimate the measurement error attached to the use of a test or instrument. When applied to psychological tests, reliability most often relates to either the internal consistency (do items in the test measure the ‘same thing’) or to temporal consistency. The latter form is referred to as ‘Test-Retest Reliability’ and is most often reported as the degree of correlation between first test and retest performance.

A number of studies have published test-retest reliabilities in healthy volunteers and patients.

- Lowe and Rabbitt (1998) assessed 4-week test-retest reliability from 162 healthy volunteers:
  - Delayed matching to sample (DMS) number correct $r=0.56$
  - Paired associates learning (PAL) first trial memory score $r=0.68$, average trials to success $r=0.86$
- Spatial working memory (SWM) total errors $r=0.68$

Leeson and colleagues (2009) assessed test-retest reliability over 1 and 2 years within 25 healthy volunteers and 104 patients with schizophrenia:
• Spatial working memory (SWM) between errors: healthy volunteer 56-week $r=0.52$, 66-week $r=0.65$; schizophrenia 74-week $r=0.62$, 100-week $r=0.62$; strategy: healthy volunteer 56-week $r=0.54$, 66-week $r=0.60$; schizophrenia 74-week $r=0.40$, 100-week $r=0.57$

• Fowler and colleagues (1995) assessed 1-month test-retest reliability in healthy volunteers, patients with questionable dementia and patients with Alzheimer’s disease:
  - Delayed matching to sample (DMS) number correct (12-second delay): healthy volunteers $r=0.52$, questionable dementia $r=0.57$, Alzheimer’s disease $r=0.84$
  - Paired associates learning (PAL) total errors: healthy volunteers $r=0.64$, questionable dementia $r=0.71$, Alzheimer’s disease $r=0.88$

Barnett and colleagues’ (2010) review outlines 3-month test-retest reliability in patients with schizophrenia:
  - Spatial working memory (SWM) between errors 12-week $r=0.65$, 14-week $r=0.83$; strategy 12-week $r=0.75$, 14-week $r=0.75$

Furthermore, Harrison and colleagues’ unpublished manuscript assessed test-retest reliability over 1-8 weeks in 100 healthy volunteers:
  - Delayed matching to sample (DMS) number correct $r=0.50$
  - Paired associates learning (PAL) stages completed $r=0.87$, total errors $r=0.68$
  - Reaction time (RTI) simple RT $r=0.54$, 5-choice RT $r=0.57$, movement time $r=0.73$
  - Rapid visual information processing (RVP) total hits $r=0.64$, latency $r=0.64$
  - Spatial working memory (SWM) between errors $r=0.70$, strategy $r=0.63$

Change in individual’s performance - measurement error or ‘real effect’?

One consequence of using a correlation coefficient as a reliability measure is that a uniform change in performance across a group will yield a high correlation coefficient, even though subjects may have improved or declined by the same rate. With psychometric measures there appears to be a high a priori expectation that performance will necessarily improve with repeated administrations of the test. Improvements in performance as a result of repeated exposure may be due to the adoption of a strategy. In circumstances such as these, changes in performance can be viewed as due to a real effect, rather than simply due to measurement error.

The foregoing discussion raises a number of issues for examining the Test-retest reliability (TRR) of the Cantab system outcome measures. The first challenge is to calculate TRR correlation coefficients for each of the outcome measures and then to determine whether retest performance improves, deteriorates, or remains the same when compared to first test scores. Changes in group performance can be determined by testing for the presence of a significant change between test and retest. However, most often one is seeking to determine whether an individual’s change in performance is due to a real effect or to measurement error.

When assessing the performance of an individual at retest, it is useful to know whether any change associated with retest performance is due to a real effect (e.g. due to physical degeneration [a worsening of performance] or due to practice or recovery [an improvement]). The standard errors of prediction reported in Table 2 can be used to determine this as they can be used to calculate a confidence interval for retest performance. If the retest performance score falls within the upper and lower bounds of this confidence interval, then retest change can be judged to be due to measurement error.

Scores beyond the confidence interval bounds are likely to be due to the influence of a real effect. This decision is subject to the usual twin statistical hazards of making type 1 and type 2 errors. When deciding whether retest change is due to measurement error or a real effect, the decision is biased toward rejecting the hypothesis that the change is due to...
measurement error. Consequently, we would suggest that only scores falling within 66% confidence intervals be interpreted as being likely to be due to measurement error.

A step-by-step procedure for determining whether an individual’s retest change is due to error or effect is described below.

**Step 1**

The vast majority of statistical techniques assume that observations are independent. Measures obtained from the same subject, even on vastly different occasions, are clearly not independent. For this reason, techniques have been devised to allow for the calculation of ‘true scores’ for which an allowance has been made for measurement error and the score actually obtained for a subject. Brophy (1986) describes the following equation:

\[ T = M + r(X - M) \]

where \( X \) is the obtained score, \( M \) is the mean of the scores on the test and \( r \) is the reliability coefficient of the test.

**Step 2**

Calculate the standard error of the ‘true’ score obtained from step 1, using the appropriate SEP from table 2 to plot the appropriate confidence interval.

**Step 3**

Determine whether the retest scores falls within or beyond the confidence interval calculated in step 2 - scores falling within the CI can be judged to be due to change consistent with measurement error. Scores beyond the CI limits are likely to be due to a real effect.

**References**

Contacting Cambridge Cognition

info@camcog.com

+44 (0)1223 810 700

**US office**
Cambridge Cognition
2750 Rasmussen Road
Park City
Utah
84098
United States

**UK office**
Cambridge Cognition
Tunbridge Court
Bottisham
Cambridge
CB25 9TU
United Kingdom