Working memory in South Africa:

Performance on a selection of computer-based neuropsychological tests

Katharine N. Fisher
Student Number: 0110238T

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF

Declaration

A dissertation submitted in fulfilment of the requirements for the degree of Master of Arts by dissertation in the field of Psychology in the Faculty of Humanities, University of the Witwatersrand, Johannesburg, May 2008

I declare that this dissertation is my own, unaided work. It has not been submitted before for any other degree or examination at this or any other university.

Signed 8 May, 2008

Katharine Nicole Fisher
0110238T

Word count: 51,197

Abstract

Working memory allows for continual updating and maintenance of information for cognitive and behavioural guidance. It provides continuity of experience and is integral to complex and adaptive human functioning. This study investigated performance on a selection of computer-based neuropsychological tests of working memory in a sample of 105 South African adults. The central aim was to examine whether demographic and computer performance variables affected performance on the computer-administered tests. Another key research question was whether commonly used tests of working memory measured domain-specific components of working memory, or tapped into domain-free executive attention. In particular, the study examined the *n*- Back Test, which had been used extensively in international research but was not sufficiently validated in the literature. An exploratory factor analysis was conducted to investigate the validity of this test.

The study found that the ability to manipulate a computer mouse affected performance particularly on the timed computerised tests, and that computer ability was also related to prior experience using a computer, confidence using a computer, gender and home language. Computer mouse ability was subsequently partialed out of the analysis as a covariate. No significant main effects of computer experience, confidence, gender or home language were found when computer mouse ability was removed from the analysis. This suggested that the demographic differences in performance found on the tests may have been informed by experience and confidence using a computer rather than reflect true differences in performance between the groups. Once computer mouse ability had been partialed from the results the 2-Back condition of the *n*-Back Test correlated significantly with the backward condition of the Digit Span Test, the forward condition of the Spatial Span Test, and part B of the Trail Making Test around the use of complex executive attention, which provided some evidence for the *n* -Back Test as a measure of the executive component of working memory. However, the *n*-Back Test did not load onto the same factor as these tests, but it appeared that the *n*-Back and Digit Span Tests factored around the type of executive resource demanded by each test.

Key Words:

Working Memory; Attention; Executive Functions; Neuropsychology; Computer-based testing; Computer Mouse Test; *n*-Back Test; Stroop Colour Word Test; Trail Making Test; Spatial Span Test; Digit Span Test

Acknowledgements

I would first like to thank my supervisor, Dr. Yvonne Broom, for her warm guidance and support over the past two years. My thanks go out to Enid Schutte for rallying the students and inspiring volunteers when I needed it most. I would like to thank Peter Fridjon for his conceptual help with the statistics for the study and Digby Ormond-Brown at the Netcare Rehabilitation Hospital for his development of the computer-based tests utilised in the study and for the use of the tests. I would also like to thank the staff of the WITS Psychology department, many of which have encouraged me at various turning points on my journey.

To my mom who has been through this process, thank you for the indispensable chats and for your unwavering support. To my dad, you have given me perspective, encouragement and have piloted me right to the end. Thank you.

Most of all to Gareth: your love, patience, support and encouragement held me together. I could not have done this without you. Thank you.

Table of Contents

Chapter One: Literature

Chapter Two: Methodology

Chapter Three: Results

Chapter Four: Discussion

Reference List

Appendices

List of Tables

List of Figures

CHAPTER ONE

Literature

1.1 Introduction

Neuropsychological testing is an important means by which the cognitive-behavioural implications of brain trauma and disease are beginning to be understood. Neuroimaging techniques have become robust detectors of the location and extent of organic damage but neuropsychological tests can better gauge impairment not evident in structural damage (McAllister, 2006a). This information is integral in a clinical context for determining type and level of cognitive impairment, for constructing tailored cognitive rehabilitation programmes, and for providing valuable information in a medico-legal context (Nadolne & Stringer, 2000). Neuropsychological testing in the South African context faces a number of important challenges. One of the current challenges, both locally and internationally, involves the use of computer-based tests (Davies, Foxcroft, Griessel, & Tredoux, 2005; Paul, Williams & Richard, 2005).

A key concern is that while information technology is designed to diminish human error, the technology employed interposes a mediating factor so that tools, such as computer-based tests that appear neutral and objective, may not always be contextually sensitive. Another concern is the low level of computer literacy in South Africa, which could potentially affect the performance of those unfamiliar with computers. However, the use of such tools has clear benefits such as reducing the time taken to administer a test, standardising the administration and instructions of a test, enhancing the precision of timing, reducing data-handling errors and allowing for adaptation of tests to the level of knowledge or skill of the test-taker (Davies et al., 2005). These benefits need be weighed up against the adverse effects of their use (Bush, Johnson-Greene, & Naugle, 2002). Provided measures are put in place to ensure that the psychometric properties of the tests are maintained, that the potential for error through misapplication and misadministration is clarified, and that level of computer experience on the part of the test-taker is accounted for, computerised testing may play a constructive role in the context of neuropsychological testing in South Africa.

The *n*-Back Test is a relatively new, computer-based neuropsychological test that has not been previously used in the South African context. It has been utilised extensively in international neuroimaging and neurobehavioural research in recent years to investigate the construct of *working memory* (Owen, McMillan, Laird & Bullmore, 2005). Though the *n*-Back Test has both content and face validity as a measure of working memory, it has not been validated sufficiently in the literature (Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005; Kane, Conway, Miura & Colflesh, 2007). To begin to investigate the validity of this measure, it is necessary to compare the *n*-Back Test with other tests that tap the cognitive constructs underpinning working memory.

The term ‘working memory’ refers to a metaphor that developed to expand upon the concept of short-term memory and its theoretical components. Working memory is defined as a limited capacity system that is involved in the temporary storage, maintenance and manipulation of information and that facilitates the processes of human thought by providing a link between the ability to engage stimuli from the environment, to store information long-term and to act according to intention (Baddeley, 2003). There are many approaches to working memory, which converge upon the concept of a limited capacity system of attention that is supported by external storage facilities, reflecting a fractionation within working memory (Repvos & Baddeley, 2006). Working memory provides a link between theoretically distinct concepts of perception, attention and memory as it illustrates how these concepts comprise a number of subsystems that interact via a process of executive control (Baddeley, 1998b).

Working memory is implicated in the breakdown of cognitive-behavioural functioning in a number of organic and psychiatric brain disorders. For example, due to the acceleration-deceleration and contact mechanisms involved in traumatic brain injury, damage is often quite predictably distributed amongst frontal, temporal and cerebellar areas of the brain, which tie into the neural circuitry of working memory (McAllister, 2006a, 2006b). Working memory deficits have also been found in patients with Alzheimer's disease (Baddeley, Bressi, Della Salla, Logie & Spinnler, 1991; Baddeley, 1996c), Parkinson's disease (Gabrieli, Singh, Stebbins & Goetz, 1996), Schizophrenia, (Conklin, Curtis, Katsanis, & Iacono, 2000), Multiple Sclerosis, (Pelosi, Geesken, Holly, Hayward & Blumhardt, 1997), Human Immunodeficiency Virus (HIV) (Chang et al., 2001) and other disorders.

Working memory has proved to be an important and challenging concept in neuroscience and modern experimental psychology (Dudai, 2002). Models of working memory are inextricably located within the progression of theory and models of short-term memory (Baddeley, 1996c). Consequently, a brief historical overview of short-term memory research and theory is necessary to contextualise and clarify the development of this concept. Alternative approaches to working memory, such as attention and capacity theories, are discussed to provide a holistic view of working memory theory and how it links with other aspects of cognition. Neuroscientific evidence for working memory is presented to reveal how perceptions of the structural and functional processes underpinning memory have evolved to accommodate new findings based on an array of neuroscientific, experimental and behavioural evidence. A working definition of working memory for the study is outlined and discussed. A selection of cognitive-neuropsychological tests operationalising working memory for the study are delineated, followed by a discussion of the advantages and disadvantages of computer-based testing in neuropsychological assessment. The review concludes with a discussion of the rationale and aims for the study.

1.2 Historical Overview

Prior to the emergence of modern cognitive psychology, the field underwent a major transformation when the concept of *information processing* was introduced (Baddeley, 1996c; Bower, 2000). The discovery of capacity limits in the human information processing system contributed significantly towards the concept of capacity-limited storage of information, from which theories of short-term memory developed. Broadbent (1958) put forward one of the most influential models of information processing based on the discovery of a temporary information store that relied on rehearsal for maintenance of information short-term. One of the most important findings was that temporary memory encoded and rehearsed information in a speech-based form (Baddeley, 1996c).

Memory had been initially conceptualised as a single system (Logie, 1995). With the introduction of the information processing perspective, evidence began to accumulate for distinct temporary and longer-term memory stores. Broadbent's concept of a distinct temporary memory store was debated. Some theorists suggested that rehearsal of information to a temporary store did not provide evidence for separate short- and long-term memory systems, but rather demonstrated that rehearsal strengthened an item so that it was better retained in a singular memory store. By the middle of the 1960s, however, evidence for distinct memory stores began to accumulate and the view of a singular memory system was largely abandoned (Baddeley, 1996c).

Following the information-processing perspective, substantial evidence accumulated to consolidate the distinction between temporary and long-term memory. Waugh and Norman (Norman, 1976) developed a theory of short-term memory founded on the concept that rehearsal not only maintained information in a temporary store, but also assisted in the *transfer* of that information to a longer-term store. Their model of memory was derived from the discovery of primacy and recency effects that occurred in serial recall of verbal information (Sternberg, 1996). Experimental tasks revealed that, when given a list of words or numbers to remember, subjects tended to remember items presented at the end (recency) and the beginning of a list (primacy) respectively, more easily than those from the middle of the list on immediate recall. However, when recall was delayed and the subject was required to recall the original list upon completion of an interference task, the recency effect dissolved but the primacy effect remained (Greene, 1992). This finding suggested that the first items on a list were rehearsed and entered into long-term memory whilst the last items on a list were retained in a shorter-term store. The recency effect occurred in the context of immediate recall only because the last items had not yet decayed from the short-term memory store. However, the introduction of an interference task prevented the last items on a list from being rehearsed and so these items could not be transferred to long-term memory (Norman, 1976).

At first, primacy and recency effects seemed to provide convincing evidence for distinct memory stores. However, the recency phenomenon was contested as evidence for the existence of temporary memory. A set of experiments demonstrated evidence of a recency effect that occurred over the long term, which was not possible according to the premise of recency (Neath & Surprenant, 2003). Another concern derived from neuropsychological evidence of patients who demonstrated normal recency effects but very poor short-term memory span, which implied that the short-term store was not the only mechanism responsible for the retention of recent information. Other experiments that attempted to correlate recency and short-term memory span discovered that the relationship between these constructs was weak in normal adults. In children, it was discovered that memory span and recency did not develop concurrently. The recency effect therefore did not provide sufficient evidence for the dissociation of short- from long-term memory. Another limitation with the model was that it only explained verbal storage and rehearsal and neglected alternative methods for short-term memory retention, such as visual imagery (Logie, 1995).

Many of the early stage models of short-term memory were criticised for oversimplifying the processes involved in memory (Logie, 1995). Shiffrin and Atkinson (1968) put forward an influential alternative that diverged from previous models. They broke away from the serial information-processing stage paradigm and proposed that memory was a more complex system comprising a number of interacting mechanisms (Hulme & Mackenzie, 1992). They hypothesised that structural components were distinct from processing of information within memory. One feature of their model that did not diverge radically from earlier models was that they maintained that information was transferred from a short-term store into long-term memory via a process of rehearsal. However, they focused on the variability in human memory in relation to the form and quantity of information that was transferred (Shiffrin & Atkinson, 1968).

Atkinson and Shiffrin advocated that memory could be divided into three structural components (Baddeley, 1996c). The first component was a sensory register, which was modality-specific so that a visual sensory register dealt with visual information, an auditory register with auditory information, and a haptic or touch register was responsible for touch information. The second structural component was the short-term store, which was limited in capacity. The third component was a long-term store, which had a substantial capacity to hold information for long periods of time and was relatively permanent (Sternberg, 1996). They also introduced the concept of control processes in memory, which were situation-specific processes that were unique to the individual, were partly under individual control, and could differ from task to task (Reynolds & Flagg, 1983). Examples of control processes included selective rehearsal of information, selective coding of information, and recovery of information from long-term memory (Neath & Surprenant, 2003). The selective nature of control processes, and the effect that different control strategies had on encoding of information to long-term memory, were highlighted.

Atkinson and Shiffrin's theory was supported by a number of significant sources. Neuropsychological evidence supported a distinction between short- and long-term memories (Reynolds & Flagg, 1983). For example, neuropsychological patients with classic amnesia could not append new information to long-term memory (Gazzaniga, Ivry & Mangun, 2002). However, their performance on immediate short-term memory tasks was relatively unimpaired. These patients had preserved ability to remember bits of information from moment to moment, but could not recall the same information after a significant delay (Banich, 2004). This finding suggested that these patients could retain information in a short-term store but that the information was not transferred to a longer-term store.

Other patients displayed the converse disorder of memory whereby remote memory and long-term learning ability was preserved but short-term memory was significantly impaired, demonstrated by the fact that patients performed poorly on auditory short-term memory tasks whilst being able to retain new information over time (Baddeley, 1996c). This finding substantiated the concept that short-term and long- term memories were independent and took this idea one step further; it demonstrated a double dissociation (Philips & Baddeley, 1989). This meant that primary and secondary memory were entirely distinct systems based on evidence from, "…contrasting patterns of memory impairment in neuropsychological patients" (Logie, 1995, p. 10). The latter phenomenon provided powerful evidence for a distinction between short- and long- term memory systems in the brain.

A paradoxical problem with Atkinson and Shiffrin's model emerged in the evidence of a double dissociation (Baddeley, 1996c). Whilst it substantiated the concept of more than one memory store, it generated a problem in terms of the structure of the model. Shiffrin and Atkinson (1968) advocated that information first had to go through a short-term store and, through the process of rehearsal, could then only be transferred to a long-term store. Neuropsychological evidence suggested, however, that short-term storage of information could be impaired but information could still somehow be encoded into long-term memory. Atkinson and Shiffrin had also proposed that short-term memory acted as a buffer or gateway to long-term memory (Flagg & Reynolds, 1983). This implied that if short-term memory were impaired, long-term memory would also be affected. Neuropsychological evidence revealed that this was not always the case. Short-term memory could be affected without any damage to long term memory stores and vice versa (Banich, 2004; Bower, 2000).

Atkinson and Shiffrin did not explain the concept of control processes in enough detail, and could not clarify how certain control processes resulted in better encoding of information to long-term memory than others (Logie, 1995). In an attempt to explain this phenomenon, a theory of 'levels-of-processing' was put forward in the literature (Baddeley, 1996c). This theory assumed that the semantic properties of information determined how 'deeply' or 'shallowly' an item was encoded into long-term memory and how easily such information could accordingly be recalled (Baddeley, 1996c). Much of the experimental research of the 1970s was influenced by the levels-of-processing approach. A problem with this approach arose in the definition of 'depth' and the criteria that determined how deeply an item was processed. Another issue was the fact that this approach dealt primarily with long-term memory, not placing enough emphasis on how storage of information was allocated to short- or long-term memory (Logie, 1995). The levels-of-processing approach also focused almost exclusively on verbal storage, failing to describe how other information was stored in the short- or long- term memory systems (Baddeley, 1996c).

1.3 Working memory

Baddeley and Hitch (Baddeley, 1982) put forward a model of ‘working memory’ that considered memory from a different angle. They proposed that memory was a more multifaceted system than initially conceived. The key point of departure of this model from previous theories was that working memory accessed both long- and short-term memory and referred to the temporary storage and activation of information from long-term memory and from the environment (Sternberg, 1996). Baddeley and Hitch proposed that short-term memory was a tripartite system comprising a system of attentional control– the *central executive* – that was supported by two distinct supplementary systems. The supplementary systems comprised a verbal system, the *phonological loop*, and a visual-spatial system, the *visuospatial sketchpad* (Baddeley, 1996c). The central executive was an overarching system that co-ordinated and organized the visual and verbal components of working memory, and was partly under individual control (Baddeley, 2000a). The supplementary systems constantly contended for central executive resources (Gazzaniga et al., 2002).

Baddeley and Hitch adopted the term ‘working memory’ to emphasise the functional aspect of short-term memory and its role as part of a system responsible for maintaining information short-term and manipulating that information whilst performing multiple cognitive tasks (Baddeley, 2003). Working memory included some of the capacity constraints of short-term memory but could access and manipulate attention and executive functioning (Zillmer & Spiers, 2001). Their model of working memory diverged from previous theories of short-term memory in three important ways. They hypothesised that memory was not a single system, but was a system made up of entirely separate parts operating together. They emphasised the role of working memory in *combined* storage and processing of information. They also proposed that this system was not limited to short-term retention of information alone, but that it operated in alliance with other aspects of cognition such as language, comprehension, and reasoning (Hancock, LaPointe, Stierwalt, Bourgeois & Zwaan, 2007).

This concept developed from experiments that attempted to mimic short-term memory deficits in normal experimental subjects (Baddeley, 1996c). Baddeley and Hitch conducted a series of experiments on the concurrent performance of subjects on a memory span measure and tasks that involved comprehension, learning or reasoning (Baddeley, 1982; 1998a). They discovered that while a short-term store was involved in memory span tasks, such as the digit span task, it was flexible enough to be involved in other aspects of cognition simultaneously. This finding led to the concept of a functional memory system that held information 'on-line' whilst the subject was involved in executing complex cognitive tasks. In effect, that short-term memory acted as a working memory system that facilitated the ability to process and manipulate information in parallel (Neath & Surprenant, 2003).

1.3.1 The Phonological Loop

One of the best-researched subcomponents of working memory was a subsidiary system, termed the *phonological loop*. Baddeley accounted for two subsystems within the phonological loop: a *phonological store* and an *articulatory control process* (Hulme & Mackenzie, 1992). The phonological store was a storage facility that held information momentarily in the form of memory traces. The articulatory control process was responsible for the rehearsal of information so that it did not decay from the phonological store (Baddeley, 2003). The phonological loop was developed following four important discoveries in research, revealed in verbal serial recall tasks. These included the phonological similarity effect, articulatory suppression, irrelevant speech and the word-length effect (Neath & Surprenant, 2003).

The first line of evidence for a phonological store that was responsible for maintaining memory traces extended from well researched and documented cases of an acoustic similarity effect that occurred when subjects were asked to recall a set of similar-sounding words or letters and performance significantly disintegrated, whilst words that did not sound the same but were similar in meaning did not interfere with memory recall. It was assumed that a process of verbal encoding of information to a phonological store, called *inner speech*, interfered with the recall of similar-sounding words due to the lack of distinguishing characteristics between similar sounding items and the resultant propensity for error as the memory traces faded from the store (Baddeley, 1996c; Logie, 1995).

For the same reason, subjects found it difficult to remember the order of items that sounded the same (Baddeley, 1982). For example, a subject that was asked to recall a string of letters such as B-D-T-C-E would be more likely to make an error in the order of recall than someone required to recall a phonemically non-similar string of letters (Neath & Surprenant, 2003). This effect was further investigated in a study of congenitally deaf children. The results revealed that the effect was not an acoustic one as congenitally deaf children who spoke well tended to sub-vocally rehearse information and demonstrated a phonemic similarity effect (Baddeley, 1982). Similar evidence was found in congenitally anarthric patients who could not articulate, but retained the ability to sub-vocally rehearse information (Baddeley, 2003). These findings suggested that the acoustic similarity effect was more a result of phonemic than acoustic similarity. Thus the effect was re-termed the *phonological similarity effect*.

A second line of evidence for the phonological store came from the *irrelevant speech effect* (Baddeley, 1996c). This effect was discovered in experiments in which subjects were divided into groups whereby they were either asked to recall items that were presented with no background distractions, or with irrelevant speech playing in the background. Subjects that were not exposed to irrelevant speech could remember the items better than those who were presented with irrelevant speech whilst concurrently trying to remember the items (Neath & Surprenant, 2003). When the irrelevant speech was phonemically similar to the items to be remembered, the effect was more pronounced (Logie, 1995). To ensure that irrelevant speech effect did not affect performance solely due to distraction, another task was conducted in which subjects were divided into groups and asked to recall items that were presented to them either with no distractions or with a tone played in the background (Baddeley, 1996c). The introduction of a tone did not affect recall. This phenomenon suggested that as irrelevant speech was speech-based, it entered directly into the phonological store and interfered with the memory traces of other items in the store.

Evidence for an articulatory control process that was responsible for sub-vocal rehearsal of items so that they could enter the phonological store came from studies that explored the effects of *articulatory suppression* on short-term memory (Baddeley, 1982). Articulatory suppression involved actively preventing a subject from subvocally rehearsing items by getting the subject to repeat the word 'the' whilst visually viewing the items concurrently. This method produced a more dominant effect than irrelevant speech (Baddeley, 1996c). Items that sounded similar were more likely to be recalled correctly than they were without articulatory suppression. It was assumed that this phenomenon occurred because articulatory suppression prevented rehearsal of the items, resulting in less phonological interference, which led to better recall of similar-sounding items (Neath & Surprenant, 2003).

Further evidence for an articulatory control process came from studies conducted on word length and its effect on memory span. This effect produced a decline in memory span that accompanied an increase in word length, termed the *word length effect* (Baddeley, 1996c). It was proposed that the length of the word determined whether it was likely to be remembered, with shorter words likely to be recalled more accurately than longer words (Baddeley, 2002). Research revealed that subjects tended to only remember words that they could pronounce within two seconds (Cowan, 1997). Further studies confirmed that the word-length effect was based on the speed at which a subject could sub-vocally articulate the words. According to this finding, memory span was dependent on the amount of time it took to articulate a word rather than the complexity of the word itself. This effect suggested that rehearsal of verbal information within the articulatory control process was time-dependant. Long words were rehearsed more slowly, which led to forgetting and to a reduced span for longer items (Baddeley, 1996c; Logie, 1995)

Further evidence for a phonological store and articulatory control process within the phonological loop came from a number of sources (Neath & Surprenant, 2003). On the one hand, the word-length and phonemic similarity effects that were revealed in speech-based encoding were eliminated by articulatory suppression when items were visually presented. This occurred because articulatory suppression prevented sub-vocal rehearsal of the items so that they were not confused with similar-sounding words or affected by the amount of time taken to pronounce the word, respectively. However, when information was presented in an auditory format the word-length effect was eliminated by suppression but the phonemic similarity effect still occurred, although in a slightly weaker form (Hulme & Mackenzie, 1992).

These findings suggested that when information was presented in an auditory form it was automatically encoded into a speech-based format, bypassing the rehearsal phase and transporting directly into a passive store. In other words, spoken material was automatically assimilated into a speech-based phonological store. Visually presented information, on the other hand, had to first be encoded into a speech-based format before entering the phonological store. This distinction also accounted for the irrelevant speech effect, as spoken material would automatically transfer into the phonological store and interfere with the rehearsal of visually presented items so that they could not adequately enter the store (Logie, 1995). However, non-speech interference would not affect sub-vocal rehearsal, as it could not access the store (Neath & Surprenant, 2003).

Neuropsychological evidence provided further substantiation for the distinction between an articulatory control process and phonological store. Certain patients demonstrated deficits that suggested a damaged phonological store (Logie, 1995). These patients tended to suffer damage to the left temporoparietal area of the brain. They did not display phonological similarity or word-length effects on recall of visually presented items on a serial recall task, suggesting that they did not use their malfunctioning phonological store. When they were presented with verbal material, however, they were forced to use the store and memory span dropped significantly (Baddeley, 2003). Conversely, dyspraxic patients that did not have an accompanying language deficit and could articulate effectively demonstrated normal phonological similarity and word-length effects, and normal use of the phonological store. However, as they could not control the processes responsible for internal ‘speech-motor programmes' they had a poor short-term memory span. This suggested an inability to sub-vocally rehearse information and a defective articulatory control process (Baddeley, 2003; Logie, 1995).

1.3.2 The Visuospatial Sketchpad

Baddeley and Hitch proposed a second subsidiary system of working memory, the visuospatial sketchpad, which was a visual-spatial version of the phonological loop. They advocated that visuospatial information was abstractly stored in long-term memory, but that the encoding, manipulation, and retrieval of such information relied on the executive constituent of working memory. Baddeley (1996c) asserted that the visuospatial sketchpad could be divided further into distinct but interrelated systems for visual and spatial information. Visual memory was distinguished from spatial memory as the "retention of static visual arrays," as opposed to the retention of dynamic spatial information or "movement through space", respectively (Logie, 1995, p.78). In other words, visual memory retained the geometric and colour properties of information, whereas spatial memory retained both physically presented and imagined movement. The model proposed that the visual imagery and spatial sub-systems were not as instant or as easy to encode as the phonological system and the sketchpad was consequently slower and more demanding on the central executive (Logie, 1995).

Evidence for a distinct visuospatial system in working memory was based on a set of dual task experiments (Baddeley, 1982; Neath & Surprenant, 2003). In these experiments subjects were required to perform more than one task at a time, under the assumption that dual task performance was possible provided the modalities required for performance were not the same. Subjects were required to complete one verbal and one spatial task. In the verbal task they could respond in one of two modalities: either in the same modality (verbal) in which they answered out loud, or in a different (spatial) modality in which they pointed to the correct answer. In the spatial task they could also respond in the same (spatial) modality or a different (verbal) modality. In the verbal task the subject was required to remember a sentence whilst simultaneously responding, and in the spatial task the subject had to retain a moving image in memory during response.

The experiments revealed that pointing interfered with the process of visual-spatial imaging. It did not, however, affect performance on the verbal memory task. The subject's response was also slower when the task and the method of response were the same. In other words, when the verbal task was accompanied by a verbal response, and when the visuospatial task was accompanied by a visuospatial response. This finding suggested that the visuospatial task interfered with visuospatial memory in the same way that phonemic similarity interfered with verbal memory, and accordingly provided evidence for distinct systems of verbal and visuospatial representation. Alternative explanations were ruled out (Neath & Surprenant, 2003).

Research did not initially support separate visual and spatial sub-systems of the sketchpad. However, cognitive, neuropsychological and neuroimaging evidence accumulated over time that did support a distinction between these subsystems in the brain (Logie, 1995). For example, a set of cognitive experiments revealed that visual imagery and spatial tracking interfered with each other in a dual task experiment, suggesting that they utilised the same modality and so could not be performed concurrently. In contrast, both the visual and spatial tasks were unaffected by verbal interference, dissociating each from verbal working memory. Neuropsychological evidence of a separate ventral visual processing stream, responsible for object information, and dorsal visual processing stream, responsible for spatial information, also supported a distinction between visual and spatial sub-systems in the brain (Postle, D'Esposito & Corkin, 2005).

Initial evidence for this distinction came from gunshot brain damaged victims from World War One. Some of these patients could locate objects but could not identify them, whereas other patients revealed the converse disorder whereby they could identify objects but could not locate them accurately (Baddeley, 1996c). However, the inability to identify objects may have reflected an inability to use language to describe an object rather than a visual memory deficit. Other neuropsychological patients revealed an inability to manipulate or remember spatial information. For example, they could not mentally rotate an image or remember a spatial route. However, these patients could still use a visual imaging process to describe objects that they had seen, thereby manipulating and recalling visual information (Baddeley, 1996b, 2003). Neuropsychological evidence demonstrated that patients with spatial deficits suffered damage to the parietal lobes whilst visual imaging impairment tended to relate mainly to damage to the occipital lobe of the brain (Baddeley, 1996b). Additional neuropsychological evidence found that certain patients could perform spatial but not pattern span tasks whilst others could perform pattern span but not spatial span tasks, providing further evidence for a distinction between visual and spatial systems in the brain (Baddeley, 2000b).

Neuroimaging evidence also converged to support a distinction between visual and spatial memory sub-systems within the visuospatial sketchpad. A meta-analysis of twenty neuroimaging studies found that spatial working memory was related to right prefrontal activation whereas visual working memory tended to activate the left hemisphere of the brain (Gazzaniga et al., 2002). An interesting finding from this meta-analysis was that visual object information stimulated bilateral activation in certain studies. On further investigation it was found that visual items could, in some cases, be recoded into a verbal format and were transformed into verbal items. This finding was important, as visual imaging had been dissociated from verbal coding of visual stimuli in Baddeley and Hitch's model of working memory. However, neuroimaging evidence suggested that visual working memory could possibly rely on a verbal coding mechanism in order to successfully encode visual information (Postle et al., 2005).

A recent study suggested a bias in the amount of central executive resources demanded by the visuospatial sketchpad (Rudkin, Pearson & Logie, 2007). First, the study demonstrated that sequential processing of visuospatial information, particularly spatial processing, placed greater demands on the central executive than simultaneous visuospatial processing. Evidence for this finding came from the fact that concurrent performance of a central executive task disturbed serial sequential spatial memory but did not affect simultaneous visuospatial processing or recall. The study additionally found that performance of random generation tasks, also a central executive task, disrupted serial spatial memory. These findings suggested that central executive tasks interfered with serial spatial recall in the same way that irrelevant speech interfered with verbal memory.

Neuropsychological evidence supplemented this finding. One study found that the Dorsolateral Prefrontal Cortex (DLPFC) was activated during temporary maintenance of the position of sequentially presented spatial stimuli. Another study demonstrated sustained activation of the DLPFC during the retention of serially and sequentially presented spatial stimuli, predominantly when capacity demands were high (Rudkin et al., 2007). Neuroimaging research also found that monitoring of location differentially activated the right DLPFC (Owen et al., 2005). Activation of the DLPFC during these tasks suggested that retention of visuospatial information in working memory was heavily dependent on an area of the brain responsible for executive control, continuous updating of information in working memory and allocation of attention across tasks (Nyberg & Cabeza, 2000). Recent research also demonstrated that the demand for executive resources in serial sequential processing of visuospatial information was greater than that for serial processing of verbal information. This suggested that the demand on central executive resources during serial sequential spatial memory tasks was not due to the demands elicited by serial recall alone (Rudkin et al., 2007).

1.3.3 The Central Executive

Baddeley and Hitch proposed a third component of working memory - the *central executive* - a system of control that supervised many aspects of cognition and, in particular, co-ordinated the verbal and visual subcomponents of working memory. The central executive component was the most important in terms of the model, but the least understood (Baddeley, 2003). The central executive was first modelled on Norman and Shallice’s theory of a Supervisory Attentional System (SAS) that coordinated attentional processes in the brain and allocated capacity-limited executive resources (Eyesenck & Keane, 1995; Gow & Stuss, 1992). Baddeley contended that an important function of the central executive would include the executive aspect of the SAS.

The concept of the central executive gained popularity once a scientific link between the frontal lobes of the brain and central executive functions was discovered. It granted credibility to the central executive component of Baddeley and Hitch's model. It was widely accepted that the Dorsolateral Prefrontal Cortex (DLPFC) played a critical role in working memory (Goldman-Rakic, 1993). Neuropsychological research demonstrated that central executive dysfunction was related to dysfunction of the frontal lobes, providing further evidence for a relationship between working memory, central executive functions, and the frontal lobes of the brain (Goldberg, 2001; Sbordone, 2000).

The psychometric approach, on the other hand, examined central executive functions from the perspective of individual differences. Psychometric research discovered that individuals had different capacities for retaining information in working memory (Baddeley, 1996c). The ability to synchronize information from different sources was identified as a crucial prerequisite for performance on working memory capacity tasks. Two key executive processes necessary for working memory capacity were identified. The first was temporary storage and activation of information (maintenance), and the second was concurrent maintenance and manipulation of information (Goldman-Rakic, 1993). The maintenance of information in working memory relied on the ability to keep information active and available, and hold it 'on-line' long enough to execute several cognitive tasks simultaneously (Baddeley, 1998b). The manipulation of information in working memory demanded the executive ability to focus attention, to divide and switch attention in order to process information in parallel, and to effectively select, control and manipulate information via a process of executive control (Banich, 2004; Goldberg, 2001; Goldman-Rakic, 1993).

Baddeley (1996c) proposed an approach to understanding central executive processes in working memory, which was to return to the concept of a 'homunculus', or hypothetical central controller in the brain. He asserted that the concept of the homunculus could be useful in propelling future research, particularly if the roles of the homunculus were fractionated and outlined in detail. This approach attempted to understand executive processes as they related to the *functions* of a homunculus or central executive system in the brain (Hazy, Frank & O'Reilly, 2006). Baddeley's concept of the functions of the central executive in working memory came from four major lines of evidence (Baddeley, 1996a).

The first function of the central executive in working memory comprised the ability to coordinate performance on two tasks, or to *divide attention* between simultaneous tasks. Evidence for this function came from studies of dual task performance in patients with Alzheimer's disease (Baddeley, 1996c). The type of memory deficit involved in Alzheimer's disease incorporated not only long-term memory impairment but working memory deficits too. Baddeley proposed that as the central executive by definition involved coordinating the slave systems of working memory, the working memory deficit that was present in Alzheimer's patients could be a result of the breakdown of the central executive system as a whole as opposed to the breakdown of either of the verbal or visuospatial systems independently.

A study was conducted in which the verbal and visual components of working memory were tested individually in a group of Alzheimer's patients and in two control groups (Baddeley, 1996a). A subsequent task, combining the slave systems, was then performed. The results confirmed that whilst the ability to perform the verbal and spatial tasks independently was relatively unimpaired in Alzheimer's patients, the ability to combine performance on the verbal and spatial tasks, and to divide attention between these tasks was significantly impaired (Baddeley, 2003). The inability to perform tasks concurrently was also significantly correlated with the progression of the disease (Baddeley, 1996c). This finding suggested that the ability to divide attention between tasks in order to combine performance comprised a key function of the central executive in working memory. The finding was strengthened by the fact that frontal lobe patients with executive impairment also demonstrated difficulty in combining performance on two tasks, and that normal elderly controls did not display a deficit in dual-task performance (Baddeley, 1996a).

The ability to *switch attention* between tasks that required constant monitoring comprised the second function of the central executive in working memory. Evidence for this function came from random generation tasks (Baddeley, 1996a). Random generation experiments required that the subject generate as many letters as possible, producing them in as random a sequence as possible, and at different rates. Evidence from these experiments converged upon the hypothesis that there was a system that was limited in capacity on which the process of random generation relied, so that the generation of letters was less random and slower as the rate of generation increased. Whilst the results were consistent across a number of studies, the findings were not suitably explained. Baddeley conducted a number of experiments testing various hypotheses to explain these findings. The first set of experiments tested the idea that random generation provided evidence for a limited capacity system of general executive functioning. Baddeley found evidence to support this concept, when a verbal memory span task and a spatial random generation task were performed concurrently (Baddeley, 1996a).

However, when required to perform two random generation tasks, each relying on the same domain, instead of being unable to perform the task subjects could cope with the tasks fairly successfully (Baddeley, 1996a). This suggested that random generation did not depend on the ability to divide attention between simultaneous tasks, which would be severely limited in the event that tasks relying on the same domain were combined, but had to rely on another executive function. Baddeley proposed that rapidly switching attention, rather than dividing attention, constituted the executive component underlying random generation. He tested the hypothesis that switching attention under a time constraint would disturb performance on random generation tasks. The study required that subjects complete a version of the Trail Making Test, which was a cognitive test that relied predominantly on the ability to switch attention between stimuli. Subjects were simultaneously required to perform a random generation task. The experiment found that concurrent switching significantly affected performance on the random generation task (Baddeley, 1996a). This finding suggested that switching attention was another separable function of the central executive.

The ability to *selectively focus attention* and *inhibit* distracting information comprised the third function of the central executive in working memory. Evidence for this function came from an experiment that demanded the ability to focus attention selectively on one set of stimuli whilst ignoring interference from another (Baddeley, 1996a). The subject groups were middle-aged and elderly subjects. The study was based on the premise that elderly subjects would have compromised central executive resources compared to middle-aged subjects. In the first trial, subjects responded to a task whilst simultaneously ignoring irrelevant stimuli in the same and in different modalities. In the second trial, subjects switched response mode between different stimuli whilst simultaneously ignoring intrusions. Response time slowed in both groups when the subject was required to ignore irrelevant stimuli across both trials. Once intelligence was partialed out, age alone determined performance in tasks that required inhibition of stimuli that were in the same modality as the task stimuli, with increased slowing of response in the elderly group. Based on the hypothesis that executive functions were disturbed in elderly subjects, this finding suggested that the ability to selectively focus on one set of stimuli while inhibiting irrelevant stimuli, was a separate function of the central executive (Baddeley, 1996a).

The ability to activate, maintain and *manipulate* information in *long-term memory* comprised the fourth function of the central executive in working memory. Evidence for this function came from working memory span tasks (Baddeley, 1996a). These tasks required that subjects listen to a set of sentences and remember the last word in each sentence. At the end of the sequence of sentences, the subject had to recall the last word from each sentence in order. Due to the interference that was caused by the sentences between each word, the information could not theoretically be stored in a short-term store. This suggested that the task relied on the ability to encode, to access and to activate the words in long-term memory. Findings from these tasks suggested that a separable function of the central executive related to the temporary activation, maintenance and manipulation of information from long-term memory (Baddeley, 1996a). This led to the proposition of an additional component to the working memory system, the episodic buffer.

1.3.4 The episodic buffer

Theories and models have to be able to account for experimental observations that challenge an existing model. Baddeley and Hitch's model (Baddeley, 2000a) was extended to explain *how* the central executive interacted with the subsystems of working memory, by including a fourth component, the *episodic buffer*, into the model. The episodic buffer was, "…a limited capacity temporary storage system that [was] capable of integrating information from a variety of sources" (Baddeley, 2000a, p.421). It was similar in theory to the concept of episodic memory but reflected a shorter-term storage facility so that it was not disrupted in patients with episodic long-term memory deficits. The buffer was more closely linked to the executive functioning of the frontal lobes of the brain and served as a multi-dimensional storage facility that linked the sub-systems of working memory with each other and helped to retrieve information from long-term memory. It was the buffer that tied working memory with the ability to direct attention to different tasks and to bind sources of information from different cognitive domains into logical episodes.

The episodic buffer was originally put forward to account for experimental observations that posed a conceptual problem to the phonological component of the model (Baddeley, 2000a; Neath & Surprenant, 2003). The model could not explain how subjects could recall visually presented verbal items from memory whilst engaging in articulatory suppression, as articulatory suppression theoretically prevented the rehearsal of this information so that it could not enter the phonological store. This would, in terms of the model, make it almost impossible to recall the information. Studies found that although memory span was affected, subjects could still retain a fair amount of information. Similarly, neuropsychological patients who could not retain information in the phonological loop, as evidenced in their memory span of one digit for auditory information, could remember up to four digits when information was presented visually. Encoding of this information into the visual (object) subsection of the sketchpad was ruled out, as the visual coding system was not able to retain items serially. Baddeley (2000a) suggested that as the central executive did not have the ability to store information, there had to be a storage system that operated when either of the subsystems was occupied so as to facilitate serial recall and to store different kinds of information.

Another issue with the phonological loop component of the working memory model was that it could not explain how sentence and prose span was typically much higher than single word or digit recall on serial recall tasks, particularly when prose or sentences were meaningfully related to one another (Baddeley, 2000a). The implication of this effect was that long-term memory was accessed so as to combine the meaningfully related information into chunks. The number of chunks determined memory span rather than the number of individual items. Baddeley (2000a) suggested that items could not be stored in the phonological loop, as the phenomenon whereby subjects had poor single item span but better sequential sentence span would not be possible. The finding suggested that there had to be a short-term storage system for chunked information that also had the ability to combine or integrate short- with long-term memory. The episodic buffer was subsequently put forward as a mechanism that stored some information, was not limited to the phonological loop, visuospatial sketchpad, central executive or long term memory systems but was domain-general and united information across the different components of memory.

Further evidence for the position that the episodic buffer combined information from different sources of information via domain-free executive processes was revealed in a series of studies that found that verbal and spatial span tasks were part of the same factor within a factor analytic study (Conway et al., 2005). This finding suggested that the slave systems utilised domain-general executive processes via the episodic buffer during working memory tasks.

1.3.5 Limitations of the working memory model

One of the limitations of Baddeley’s (2003) model of working memory was that the concept that items in short-term memory decayed unless rehearsed, particularly within the articulatory loop, was not entirely supported (Neath & Surprenant, 2003). The word-length effect had previously been cited as a prime example of this effect, so that longer words could not be adequately rehearsed due to a time restriction within working memory, which led to a decay of longer words from the phonological store. However, a set of experiments was conducted whereby longer words did not lead to a time-related decay and, conversely, led to an increase in recall. This finding suggested that other factors, such as the meaning of the word, could perhaps better determine whether a word was encoded and recalled. Baddeley proposed, however, that the levels of processing model could be integrated with the working memory model to explain this kind of phenomenon (Sternberg, 1996).

A second issue came from memory span tasks. In terms of the model, the length of the sequence determined how much information could be stored short-term and the capacity of the short-term store was limited to five items at any point in time. However, there was a conflicting argument that memory span data may not reflect capacity, but rather revealed the effects of serial order on short-term memory. Neath and Surprenant (2003) claimed that if the episodic buffer was involved in supporting memory span tasks, and memory span remained restricted to only a few items, the episodic buffer would necessarily be too limited in capacity to be helpful in other capacity-type tasks. Baddeley and Hitch's model was criticised for not explaining the effects of serial order and the manner in which temporal order was retained in memory. Neuroimaging research found that the area of the brain utilised in tasks that tapped into working memory was also utilised in memory for serial order, which suggested that the executive functions involved in working memory tasks were also involved in retaining serial order information (Gazzaniga et al., 2002). Baddeley's model may therefore not be as conceptually far removed in accounting for serial order effects as assumed. However, this effect needs to be fully explained with regards to the model.

The concept of interference in verbal working memory was also criticised in the literature due to a lack of evidence to support the contention that interference was primarily related to phonological similarity within the phonological loop, and due to the finding that tones could interfere with verbal serial recall (Neath & Surprenant, 2003). The model did not explain how recency effects were more effective with auditory items, which implied that acoustic information was more effectively stored than visual information. The model was also criticised for failing to describe in detail the relationship between the subsystems of working memory and long-term memory. Finally, Baddeley and Hitch's model proposed the existence of a capacity limit in the central executive but, for many years, no task could effectively measure these limits. However, recent research has begun to investigate the concept of capacity limits in central executive functioning using cognitive tasks that vary the amount of information presented and measuring the point at which performance begins to decline on these tasks (Neath & Surprenant, 2003).

Alternate approaches to working memory, following Baddeley and Hitch's original model, are diverse. However, they may not be as disparate as they may seem. Some focused on the function of attentional control in working memory, and others remained focused on explaining working memory findings in terms of original models of long-term memory (Baddeley, 2000b). Another prominent approach focused on the capacity limits of working memory (Neath & Surprenant, 2003). The main discrepancy between Baddeley and Hitch's model and other models corresponds to a difference in degree of emphasis on each of the theoretical elements and a difference in terms of the comprehensiveness of each model, as opposed to direct conflict between the models (Baddeley, 2003). Some areas of incongruity are highlighted below.

1.4 Attention and Working Memory

Phillips and Baddeley (1989) noted that research on memory, attention and perception was likely to become increasingly interconnected. Concepts of memory, perception and attention were studied and researched separately for a long time. However, under the metaphor of working memory, these concepts were linked. Dudai (2002) contended that perception of a stimulus occurred when attention allowed for orientation towards the stimulus, at which point the central executive component of working memory took over and voluntarily deduced whether the stimulus was important and deserved the primary focus of attention, additional processing and action for further use. The interaction between perception, attention, and memory, also known as 'working attention', occurred almost instantaneously. Cowan (1997) recognised that memory and attention were not one singular concept, but were intersecting and interrelated concepts. Theories of attention in working memory did not detract from Baddeley and Hitch's model but rather emphasised the executive aspect of working memory (Baddeley, 2003).

Baddeley (1996c; 1998a) contended that theories of attention should be incorporated into an understanding of the processes involved in working memory. Baddeley and Eyesenck (Eyesenck & Keane, 1995) proposed the existence of a central executive or attentional controller at the top of a hierarchical configuration that co-ordinated and controlled behaviour, and somewhat autonomous mechanisms that functioned at ground level. Within this view, the executive functions were responsible for the top-down systemic allocation of attention to various tasks. Evidence for functionally distinct components of attention, discovered in brain imaging research, included the ability to sustain, focus-execute and shift attention (Mirsky et al., 1991). These basic components corresponded closely to the functions of the central executive which included the ability to selectively focus attention by inhibiting irrelevant information, and to divide and switch attention between tasks that required constant monitoring (Baddeley, 1996a)

Cowan (1997) proposed, via an *embedded processes model*, a different understanding of the concepts of attention and working memory. The first major divergence was that working memory was comprised exclusively of processes rather than structural components. The second was that working memory referred only to information that was activated in long-term memory (Neath & Surprenant, 2003). Cowan proposed that attention activated portions of long-term memory at which point information was further activated through the focus of attention so that it could be manipulated in working memory (Cowan, 1997).

Cowan (1997) retained the idea that a sensory store maintained information for a very short period of time, but asserted that from this point information could go through one or both of two processes: sensory and/or semantic. Sensory information would remain in a short-term storage facility whereas semantic information could activate portions of long-term memory related to that information (Neath & Surprenant, 2003). Cowan concluded that without attention, sensory memory would temporarily encode information, but this information was quickly lost unless attention was directed towards it. He adopted the concept that certain aspects of attention could be ‘automatised’ for example by habituating a stimulus, and free up attentional resources for other tasks. This accounted for the ability to process information in parallel (Cowan, 1997).

One of the problems with Cowan’s model was that the concept of ‘activation’ of information was not defined in detail (Neath & Surprenant, 2003). A second issue was that the concept that information was deactivated unless reactivated within a certain time frame was not supported in the literature. The theory also could not adequately account for the retention and recall of serial order items. However, it integrated with Baddeley and Hitch’s theory of working memory, as it described the executive allocation of attention to information and how attention activated portions of long-term memory. The theory thus contributed to an understanding of the role of attention of working memory.

A recent *time-based resource- sharing model* of working memory emphasised the capacity-limits of working memory and explained these limits in terms of attentional constraints. According to this theory of attention in working memory, the time during which a task utilised attention was the defining element of cognitive load (Barrouillet, Bernardin, Portrat, Vergauwe & Camos, 2007). In terms of the model, working memory was reliant on a common attentional supply that could be utilised for executive processes, but that was limited to the relative amount of time that a task demanded attention. Thus, the current utilisation of attention inhibited the ability for other tasks to access the attentional resource. The model was put forward as an answer to the problem of a ‘trade-off’ between processing and storage in Baddeley and Hitch’s working memory model. In other words, the problem of accounting for the decrease in performance when cognitive load was increased, and the decay of information from short-term memory as a result of increased processing demand.

Four key assumptions of the model defined the role of attention in working memory. First, the model contended that a limited attentional resource was available on which processing and maintenance of information were both dependent. Barrouillet et al.

(2007) noted that this resource was otherwise known as ‘controlled attention’ or as a central executive directed attentional resource. Secondly, the model assumed that attention acted as a bottleneck so that only one kind of task, processing or maintenance, could occur at any one point in time. The third assumption was that once an item had finished utilising attention, and attention switched to another task, the original item decayed as time progressed unless the item was reactivated through focused attention and retrieval (rather than rehearsal) of the original item. The final assumption was that attention could be ‘divided’ between processing and maintenance, not by actually dividing attention but by rapidly alternating or switching between these tasks (Barrouillet et al., 2007). This theory contributed towards an understanding of how working memory tasks created cognitive load, by occupying a limited attentional resource for a period of time so that it was difficult to direct attention towards other tasks. It also explained how subjects could perform on parallel tasks by rapidly switching attention between the tasks. However, the model was criticised for focusing only on the concept of cognitive load in its account of working memory (Barrouillet et al., 2007). The comprehensiveness of the model was therefore limited. The model was useful, however, in that it provided a convincing argument for the role of attention in cognitive load.

One of the key contributions of theories of attention in working memory was the concept that the executive component of working memory could be fractionated into a number of attentional processes. Maintenance of information could be dissociated from manipulation of information in working memory. Maintenance referred to fairly simple processing of information, was achieved through simple attentional processes such as focused attention and the ability to ignore interference, and was mainly affected by storage as opposed to executive limitations. Manipulation referred to the processing of new information whilst concurrently maintaining information short-term, and relied more heavily on the executive attentional processes of working memory (Baddeley, 1996c). The relationship between executive attention and storage of information, and the trade-off that occurred as a result of this relationship, gave rise to the concept of capacity limits in working memory.

1.5 Working Memory Capacity

Capacity constraints have been investigated with regard to the physiological basis of working memory in the brain (Callicott et al, 1999). Working memory capacity has been theoretically distinguished from the concept of short-term memory capacity. Short-term memory capacity referred to simple serial memory span and short-term maintenance of information. Working memory capacity, on the other hand, has been defined as "…the extent to which a person can control and sustain the focus of attention in the face of interference and distraction" (Neath & Surprenant, 2003, p. 83). Working memory capacity included an extra role in that it involved the ability to concurrently shift attention towards processing of new information whilst attending to the task at hand (Conway et al., 2005).

Capacity limits referred to the decline in performance that accompanied an increase in processing or cognitive load placed on working memory. Functional neuroimaging studies have attempted to reveal the neurological underpinnings of the capacity constraints of working memory within the larger network of working memory in the brain (Callicott et al., 1999). Neuroimaging evidence supporting the concept of capacity limits in working memory has come from studies that have demonstrated an increase in activity in the Dorsolateral Prefrontal Cortex (DLPFC) in working memory tasks, which paralleled an increase in cognitive load. However, activity decreased in the DLPFC as cognitive load increased further, suggesting that as working memory performance declined behaviourally, so activity in the area of the brain associated with executive functions decreased (Carpenter, Just & Reichle, 2000). Capacity theories contributed to working memory theory and research by highlighting the role of individual response strategies and individual variability in working memory performance.

1.6 Neurological basis of Working Memory

The introduction of lesion studies, animal studies, single cell recording, evoked potential and functional neuroimaging methods such as electrophysiological, Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), and optical imaging into the neurological underpinnings of memory has allowed the field of memory research to expand and diversify at an enormous rate. Functional neuroimaging methods have revolutionized the field of neuroscience by providing a means of investigating the human brain as it worked (Nyberg & Cabeza, 2000). These methods involved high-definition brain scanning, which constructed a three-dimensional image of the brain as it was involved in cognitive tasks. They differed from traditional static imaging methods as they did not take a ‘snapshot’ of the brain at one point in time but rather tracked changes in the brain as they occurred (Dudai, 2002).

PET and fMRI utilised haemodynamic methods to compute the variability in cerebral blood flow that reflected neuronal activity in the brain (Nyberg & Cabeza, 2000). The method used to assess whether neuronal activity reflected real change in cognitive function first involved measuring the activity of cerebral blood flow whilst the subject completed a particular cognitive task. The second step was to measure the activity of cerebral blood flow whilst the subject completed a matched task that did not include the cognitive element under investigation. The blood flow patterns were then contrasted to assess which areas of the brain were differentially activated by the cognitive task (Nyberg & Cabeza, 2000).

Neuroimaging methods have been used to investigate the subcomponents of working memory and the areas of the brain activated by each component. Baddeley (2003) cited evidence for localisation of the basic subcomponents of the working memory system in lesion studies and neuroimaging research. Neuroimaging studies supported the finding that, generally, *verbal* working memory tasks utilised the lateral prefrontal cortex with emphasis on ventral areas anterior to Broca’s area in the brain (Gazzaniga et al., 2002). The phonological loop was the most clearly differentiated system in which the phonological store consistently activated the left posterior parietal, opercular and premotor frontal regions of the brain with Broca’s area involved in articulatory rehearsal of verbal information (Nyberg & Cabeza, 2000).

On the other hand, *visuospatial* working memory tasks generally activated occipital-parietal, pre-frontal and frontal lobes, predominantly in the right hemisphere of the brain (Baddeley, 1996c; Gazzaniga et al., 2002). Single-cell recording and neuroimaging evidence supported the dissociation between visual (object) and spatial components of the visuospatial sketchpad (Nyberg & Cabeza, 2000). First, the maintenance of visual and spatial information tended to activate ventral and dorsal pathways of the brain, respectively. Second, visual (object) information activated occipitotemporal and inferior parietal areas of the brain in the left hemisphere, whilst spatial information activated occipitoparietal and superior prefrontal areas mainly in the right hemisphere. fMRI research into the areas of the brain involved in spatial location and visual pattern recognition tasks such as the *n*-Back task also demonstrated dissociation between visual and spatial working memory. Spatial location tasks activated the parietal lobe whereas visual pattern recognition tasks tended to activate the inferior temporal lobes (Carpenter et al., 2000). The evidence converged to support the dissociation between visual and spatial subsystems in the visuospatial sketchpad.

Further evidence for a distinction between verbal, visual and spatial components of working memory has been demonstrated in a meta-analysis of 24 fMRI studies that compared areas of the brain that were activated during verbal, visual and spatial working memory tasks (Owen et al., 2005). Tasks that required identity monitoring of stimuli revealed that verbal identity-monitoring tasks utilised the left ventrolateral prefrontal cortex, the medial and bilateral premotor cortex, the bilateral medial posterior parietal cortex and the thalamus whereas visual identity monitoring tasks activated the frontal pole and dorsal cingulate areas of the brain. Monitoring of location, on the other hand, activated the right DLPFC and lateral premotor cortex and the right medial posterior parietal cortex (Owen et al., 2005). These findings provided further evidence for the dissociation of verbal and visual identity monitoring, and spatial location monitoring in the brain.

Although there was a wealth of research that attempted to localise the executive functions of working memory, this system could not be located specifically in one area of the brain. There was, however, evidence to suggest that portions of the frontal lobes were involved in different aspects of executive functioning (Baddeley, 1996a; Banich, 2004; Goldberg, 2001; Sbordone, 2000). Neuroimaging research revealed that the DLPFC was activated in tasks that demanded executive processes involved in monitoring of information, ‘temporal tagging’ and identification of items in relation to when they appeared in a sequence, continuous updating of information in working memory and allocation of attention across tasks. Executive processes including problem solving, planning and reasoning also activated the DLPFC (Nyberg & Cabeza, 2000). The DLPFC therefore clearly played an important role in the executive processes involved in working memory.

Goldman-Rakic (1993) discovered that in cellular processes in primate research, the DLPFC was distinctly involved in executive processing of information in working memory. Further animal research revealed that working memory depended on two key executive processes. The first was a process whereby stored information was maintained, and the second was a process whereby retrieved information was activated and manipulated (Gazzaniga et al., 2002). Animal studies of recordings in the lateral prefrontal cortex during visual location delayed-response tasks revealed that the cells of the prefrontal cortex remained active following the presentation of a stimulus and in the subsequent delay between presentation and recall when the animal could not see the item. Certain cells in the prefrontal cortex remained active during the whole task whilst other cells only became active up to a minute during the delay. This finding suggested that certain cells of the prefrontal cortex could be correlated with the ability to retain a memory trace and others could be correlated with the ability to manipulate information in working memory following the disappearance of an item from view (Gazzaniga et al., 2002). However, this finding needed replication in human subjects.

Neuroimaging research in human subjects revealed a reduction in the amount of activity in the DLPFC coincident with a decrease in performance on capacity-demanding working memory tasks. Callicott et al. (1999) examined the effect of capacity limits on brain activation in event potential research. They conducted a study on nine neurologically normal subjects whereby the subjects completed a task that progressively increased cognitive load and subsequently increased the demand placed on working memory capacity. Their results supported the finding that the DLPFC was activated while an item was held over a delay and while that information was manipulated. In other words, the area of the brain involved in executive functioning was also involved in tasks that manipulated the capacity of working memory. On the other hand, the difficulty of the task and the mental effort required in keeping up with the task activated a different area of the brain – the anterior cingulate (Callicott et al., 1999). In other words, the neural effect of working memory capacity was distinguished from the effects of task difficulty in the brain (Barch, Braver, Nystrom, Forman, Noll & Cohen, 1997).

Another fMRI experiment examined the maintenance versus manipulation of information in working memory via administration of a letter-span task (Carpenter, Just & Reichle, 2000). The experiment required subjects to maintain a series of letters in a particular order and then to reorganise them in alphabetical order. The study hypothesised that maintaining the letters in the same order would require simple maintenance of information whilst reorganising the letters in alphabetical order would require active manipulation of the information in working memory. The results supported previous findings that the DLPFC was involved in the active manipulation of information in working memory. However, both the DLPFC and the ventrolateral prefrontal cortex were activated during both the maintenance task and during the manipulation task. The only difference was in the degree and extent to which these areas were activated by the different tasks. The finding suggested that the same prefrontal regions of the brain could be differentially activated according to the degree to which a task demanded executive processes. It also suggested that the ability to maintain or manipulate information in working memory could rely on different amounts of executive resources rather than on entirely distinct processes in the brain.

Baddeley (2000a) speculated that the episodic buffer could not be localised to one area of the brain but could involve frontal lobe functioning, particularly with regards to *integration* of tasks in working memory. One study that investigated integration versus non-integration of verbal and spatial information under fMRI revealed that the right frontal area of the brain was more involved in the recall of integrated verbal and spatial information whereas recall of information that had not been integrated activated the posterior areas of the brain related to verbal and spatial working memory, respectively (Baddeley, 2000a). This finding suggested that the episodic buffer involved domain-general executive processes and so activated the same area of the brain when combining verbal and spatial tasks whilst individual completion of verbal and spatial tasks activated their respective domain-specific areas of the brain.

Neuroimaging research has primarily focused on localising different working memory components and processes to specific areas of the brain. However, an alternative view has suggested that cognitive processes may be the emergent properties of networks that are distributed across the brain (Carpenter et al., 2000; McEvoy, Smith & Gevins, 1998). This view suggested that differential activation of regions of the brain was determined by the amount and type of cognitive load demanded by various cognitive tasks (McEvoy et al., 1998). It also suggested that areas of the brain integrated with one another in different ways according to the particular requirements of a task and according to the cognitive strategies and skills that were developed within a particular culture (Carpenter et al., 2000). The latter concept was important in that it recognised that these networks were configured over time and they were organised and informed by cultural learning.

In summary, neuroimaging evidence suggested that working memory could be consciously accessed by the frontal lobes of the brain (Goldberg, 2001). It also suggested that the working memory system was reliant on attention, the integrity of the frontal lobes, and the integrity of the sub-cortical circuits of the brain (McAllister, 2006b). Recent evidence suggested that working memory performance was dependent on combinations of brain regions according to the amount of cognitive load and particular requirements of the task, and that these networks were informed by the cognitive skills developed within a culture (Carpenter et al., 2000). Neurobiological research has not always integrated with cognitive or experimental psychology but, in recent years, these domains have begun to inform one another (Cowan, 1997). The integration of these methodologies has become increasingly apparent, as revealed in the wide variety of research, highlighted above.

1.7 Working definition of Working Memory

The working definition of working memory is important, as it has informed the way in which the concept has been manipulated theoretically and methodologically (Tulving, 2000). The definition for working memory adopted in this study was primarily based on Baddeley and Hitch's theory of working memory, and highlighted the verbal, visual and spatial subcomponents and executive processes within the working memory system. Baddeley and Hitch's model was selected for a number of reasons. First, it has remained one of the most influential, durable, and adaptable models of working memory to date (Baddeley, 2000a, 2002, 2003; Hancock et al., 2007; Neath & Surprenant, 2003; Wager & Smith, 2003). Second, the model has been supported by a wealth of neuroimaging, lesion study, and animal research, particularly in relation to the dissociation of verbal, visual and spatial subsystems in the brain and to the executive processes that have allowed for maintenance and manipulation of information in working memory (Gazzaniga et al., 2002).

A number of concepts utilised in this research are defined and clarified. First, working memory was acknowledged as a multi-component system consisting of domain-specific subcomponents, as discovered in neuropsychological evidence and neuroimaging research. Second, it was acknowledged that a domain-free system of complex executive attention was responsible for sequencing and processing information, and selectively allocating attention between tasks. The executive system was not exclusive to working memory. It was assumed to be involved in numerous cognitive abilities and was responsible for goal-directed attention for action (Styles, 1997). Third, two key processes were identified within working memory: maintenance and manipulation of information. Maintenance of information relied on the ability to store and to rehearse information mainly within each of the domain-specific slave systems of working memory. Manipulation of information relied on the ability to access the domain-free executive attentional system, to utilise higher-order attentional abilities, and to integrate information across a number of tasks via the episodic buffer.

Short-term memory capacity was reliant on maintenance of information within each of the components of working memory, as evidenced in simple span tasks whereby verbal or visuospatial information was presented and recalled serially. It referred mainly to domain-specific storage of information within each of the subcomponents of working memory, and relied on focused attention. Working memory capacity utilised domain-general complex executive attention for manipulation of information and referred to the ability to retain information short-term whilst simultaneously switching attention to another task (Conway et al., 2005).

One means by which relationships between the components and processes of working memory have been investigated has been to examine the relationships between some of the tests currently utilised in neuropsychological assessment to assess these concepts. Cognitive-neuropsychological tests have been used as an effective means of operationalising cognitive constructs in both behavioural and neuroimaging research. Operationalisation of a construct had to be based on theory, so that the tests measured what they claimed to measure in an unbiased, fair, valid and reliable manner. Some of the tests used to operationalise different aspects of working memory for the study are outlined below.

1.8 Operationalising Working Memory

1.8.1 Cognitive- neuropsychological tests

Neuropsychological assessment of working memory traditionally distinguished measures of the components of working memory from measures of the processes of working memory (Conway et al., 2005). The current study utilised a battery including tests that had been used to investigate the domain-specific components of Baddeley's framework that each employed different processing demands on different phases of the tests. The phonological loop was explored using an adapted computer-based version of the Wechsler Adult Intelligence Scale (WAIS-III) Digit Span Test and the spatial aspect of the visuospatial sketchpad was explored using an adapted computer-based version of the WAIS-III Spatial Span Test (Quinette, Guillery, Desgranges, de la Sayette, Viader & Eustache, 2003). A visual computer-based version of the *n*-Back

Test was used to explore the visual (object) subcomponent of the sketchpad. The current study also administered tests of executive attention that did not have a storage component, but placed heavy demands on executive processes including the ability to focus, divide and switch attention, such as the Stroop and Trail Making Tests. It was proposed that the relationships within and between these tests would demonstrate how working memory was broken into different levels of storage and processing, and how these levels operated together.

n-Back Test

The *n*-Back Test was highlighted in this study, as it was one of the few tests of working memory that had been extensively utilised in fMRI research, but had not been sufficiently investigated in the literature (Owen et al., 2005). It was a relatively new test that measured working memory performance as processing load was increased (McAllister, 2006b). The *n*-Back Test had been used in conjunction with fMRI research to investigate the influence of different capacity loads on working memory (Perlstein, Carter, Noll & Cohen, 2001). Capacity load referred to an increase in the processing demand placed on working memory and a subsequent decrease in functional performance on the task (Callicott et al., 1999). Cognitive performance on this task was mirrored by a shift in neurophysiological response in the DLPFC and other areas such as the premotor cortex, the thalamus and the superior parietal lobe (Gazzaniga et al., 2002). Capacity limits on the *n*-Back Test were thus evident in the physiological network of working memory in the brain, as well as being behaviourally evident on this task (Callicott et al., 1999).

The *n*-Back Test requires subjects to constantly monitor a series of verbal or non-verbal items presented in sequence, and determine whether the item currently viewed is the same item as one presented *n* items ago (Owen et al., 2005). Conway et al. (2005) noted that *n* typically varied from 1 to 4. The subject is required to continuously update new items, whilst simultaneously recalling the last few items in the set, and then selectively disregard previous items when they are no longer of use to the task at hand. Stimulus encoding and response demands are constant across the various conditions and the only factor that varies across the tasks is information load

(Perlstein et al., 2001). Accuracy of response is measured as processing load increases.

The 0-Back condition involves no memory component and only demands response to a pre-identified item whenever it appears in a continuous sequence of stimuli. The 1-Back, 2-Back and 3-Back conditions involve the ability to monitor, revise and manipulate information to greater degrees as load is varied. In other words, the capacity of working memory is measured as the amount of cognitive load increases. Whilst 3-back conditions have been utilised in a number of studies the validity of such results have been queried, as performance on the 3-back tended to decrease significantly (Owen et al., 2005). This study included a 3-back condition in order to test the limits of performance on this task. Owen et al. (2005) noted that there were different types of *n*-Back designs, including visual, spatial, auditory and olfactory, which placed demands on completely different processing systems within the brain. A visual version of the *n*-Back Test was utilised in order to include a visual measure of working memory ability in this study.

The *n*- Back Test demanded simultaneous maintenance and manipulation of information, and the ability to selectively focus attention and ignore distracting items that had appeared previously in a sequence. The task was similar to working memory span tasks, such as reading span and operation span tasks, as it required active monitoring, maintenance and manipulation of information in working memory. It consequently demonstrated face validity as a task that tapped into working memory (Owen et al., 2005). However, a limited amount of research had found evidence both for and against the *n*- Back Test as an effective measure of working memory (Kane et al., 2007).

Digit Span Test

The Digit Span Test was a popular test that has been used to assess the capacity limits of short-term memory. The forwards component of the Digit Span Test requires subjects to maintain a sequence of numbers in memory and repeat them in the same sequence on immediate recall. It relies on the ability to selectively focus attention and maintain information temporarily in short-term memory (Lezak, 1995). Baddeley (1996c) noted that the forwards phase tapped mainly into short-term storage rather than executive abilities. Short-term serial memory span tasks such as the forwards phase of the Digit Span and Word Span Tests did not have the defining characteristic of a working memory span task: simultaneous processing of an additional task during the presentation of items, so that rehearsal of the original item was restricted (Conway et al., 2005).

The backwards phase of the Digit Span Test requires subjects to maintain a sequence of numbers in memory and reverse them on immediate recall. It demands the ability to concurrently maintain and manipulate information by switching attention between tasks so that the defining characteristic for working memory capacity is met (Conklin et al., 2000). It involves both short-term maintenance of information and the executive ability to rapidly switch attention between items and demands an, "… active manipulation of digits in the working memory buffer" (Groth-Marnat et al., 2000, p. 164). The backwards phase of the Digit Span Test was theoretically distinguished from the forwards phase in the literature (Ponsford, 2000). A number of authors proposed that the cognitive requirements for performance on the forwards phase were handled predominantly by the phonological loop whereas the backwards phase required more complex involvement of the central executive (Wilde, Strauss & Tulsky, 2004).

Evidence both for and against the backwards phase of the Digit Span Test as an effective measure of working memory capacity has been found. One study demonstrated that the backwards task tended to factor with simple span tasks as opposed to factoring with other working memory span tasks (e.g. reading span, operation span) in a factor analytic study (Engle et al., 1999). However, reading span and operation span tasks combined tasks across different domains whereas the Digit Span Test was domain-specific. It has been demonstrated that domain integration activated a different part of the brain- the right frontal area - as opposed to non-integrated tasks that activated their respective domain-specific areas of the brain (Baddeley, 2000a). This finding could be interpreted as reflecting the stronger correlation of domain- integrated tasks with each other. On the other hand, task integration could be a stronger indicator of working memory performance. A later study revealed that working memory span tasks correlated with each other only slightly better than they correlated with digits backwards. It was suggested that each of these tests might reflect a single construct with each task revealing different points on a continuum (Conway et al., 2005).

Spatial Span Test

The Spatial Span Test is similar to the Digit Span Test. However, it substitutes visual-spatial for verbal information. It requires the ability to maintain a sequence of spatial stimuli in memory and recall them in the same sequence on immediate recall and has been described as a measure of visuospatial short-term memory span (De Lillo, 2004). The requirements for the Spatial Span Test were equivalent to Corsi's Block Tapping Test (Baddeley, 1996c; Logie, 1995). In all other respects the Spatial Span Test was administered in the same manner as the Digit Span Test. For example, the Spatial Span Test includes the same number of items in a sequence, the same order of sequences, and the same number of attempts per sequence as the Digit Span Test. It also includes both a forwards and backwards phase on the test. However, this measure is truly spatial as the subject has to remember the successive location, or movement, of a set of stimuli.

Recent neuroimaging research had found that maintenance of serial sequential spatial information consistently activated the DLPFC, an area of the brain responsible for executive attention, and that processing of serial sequential spatial information demanded more executive resources than processing of serial verbal information (Rudkin et al., 2007). However, the technical manual of the Wechsler Adult Intelligence Scale (WAIS-III) proposed that the Digit Span and Spatial Span Tests were equivalent tests that each equally measured the ability to hold a verbal and visuospatial sequence of events, respectively, in working memory. Another assumption of the WAIS-III Spatial Span Test was that the backwards condition was more demanding than the forwards condition of the test. The Spatial Span Test was also assumed to be a valid test of the capacity of the visuospatial subsystem of working memory (Wilde & Strauss, 2002).

Wilde et al. (2004) re-examined the WAIS-III and Working Memory Scale (WMS-III) standardisation and clinical group data. They found some significant differences between the Digit Span and Spatial Span Tests. First, they confirmed that forward scores were, overall, significantly better than backwards scores on both the Digit Span and Spatial Span Tests in the standardisation sample of 1250 individuals. Second, they examined the percentage of subjects who performed better, equal to, and worse, on the forwards condition of each test compared to performance on the backwards conditions of each test. The analysis revealed that 92.9% of the subjects performed better on the forwards than the backwards condition of the Digit Span Test, whereas only 65.5% of the subjects had performed better on the forwards than the backwards condition of the Spatial Span Test. In addition, 2.6% of the subjects obtained better scores on the backwards than the forwards condition of the Digit Span Test, whereas 17.7% of the subjects obtained better scores on the backwards than the forwards condition of the Spatial Span Test. The remaining subjects performed at least as well on the backwards as on the forwards conditions of the Digit Span and Spatial Span Tests (Wilde et al., 2004).

Wilde et al. (2004) proposed a number of explanations for this finding. First, they contended that the Digit Span and Spatial Span tasks were not as methodologically isomorphic as assumed. They pointed out that the sequences given in the forwards and backwards conditions of the Digit Span Test were different whereas the same sequences were given on both the forwards and backwards conditions of the Spatial Span Test. They posited a practice effect; as the backwards condition followed the forwards condition of each test subjects saw each sequence for a second time, which may have aided backward span performance (Wilde et al., 2004).

A second explanation for this finding was that the Digit Span Test required subjects to remember both the individual digits and the order in which they were presented whereas the Spatial Span Test only required that subjects recall the order or sequence of presentation. Wilde et al. (2004) cited previous research that had found that recall of only the order of items, in any modality, resulted in similar or equal forward and backward spans. Their hypothesised reason for the difference in performance on the

Digit Span and Spatial Span Tests was that the backwards phase of the Digit Span Test involved the use of further executive resources for active manipulation and reversal of the digits whereas the backwards phase of the Spatial Span Test did not. Wilde et al. (2004) proposed that the Spatial Span Test relied on recall of the relative position of the items as opposed to actual storage of the items in working memory. The authors concluded that the Digit Span and Spatial Span tasks were not methodologically equivalent and could not be directly compared.

A backwards phase on the Spatial Span Test was included in this research to investigate the assumption that the backwards condition involved additional executive resources during the process of reversing a sequence of spatial items in working memory (Neath & Surprenant, 2003). However, different sequences were used on the forwards and backwards phases to ensure that backwards performance was not facilitated by repetition. A composite score summing performance on the forwards and backwards conditions of the Spatial Span Test was deemed redundant, as it was important to assess whether the forwards and backwards conditions of the Spatial Span Test measured the same or different cognitive constructs.

The Digit Span and Spatial Span Tests were selected to measure working memory in place of other working memory tests such as reading span or operation span for four main reasons. First, so that direct comparisons between verbal and spatial working memory capacity could be made according to the assumption that the Digit Span and Spatial Span Tests were analogous tests that were presented in different domains. The second reason was to test the assumption that these tests measured different aspects of working memory on the forwards and backwards conditions of each test in a single domain. The study could not subsequently be complicated by domain integration. The third reason was to test the hypothesis that the backwards phase of each test would correlate, based on the assumption that working memory capacity by definition utilised the same domain-free executive functions (Conway et al., 2005). The fourth reason was that these tests were well documented and validated tests that had been used extensively in cognitive research to examine the capacity of short-term and working memory (Groth-Marnat, Gallagher, Hale, & Kaplan, 2000; Lezak, 1995). Additionally, the Digit Span and Spatial Span Tests loaded onto the working memory indices of the WAIS III and Wechsler Memory Scales (WMS –III), respectively (Ponsford, 2000).

Executive attention

The Trail Making Test (TMT) from the Halstead-Reiten neuropsychological test battery is a well- known and well-used test that is assumed to place few demands on memory but engages the executive ability to rapidly switch attention between tasks (Baddeley, 1996a). Part A of the test requires subjects to identify a set of numbers scattered across a page in order and at speed. Part B of the test requires subjects to switch attention between identifying numbers and identifying letters of the alphabet alternately and at speed. The TMT measured "…simple cognitive processing speed…" in Part A, whilst Part B of the test measured processing speed, as well as the ability to effectively switch attention between tasks in parallel (Ponsford, 2000, p.237). Part A did not tap into any cognitive ability besides processing speed, whilst Part B relied heavily on attentional switching, a function of complex executive attention (Baddeley, 1996a). The WAIS-III and WMS- III working memory indices demonstrated significant correlations with Part B of the TMT (Ponsford, 2000). This substantiated the utilisation of the TMT as a test of complex executive attention in the assessment of working memory.

The Stroop Colour Word Test is another well-known and well-used test that does not have a memory component but demands the executive ability to focus attention and ignore distractions, and the ability to rapidly switch attention between tasks (Lezak, 1995; Ponsford, 2000). The format of the Stroop task used by different researchers was not uniform. There were many versions of this test, which varied according to the number of trials, the number of items, the number of colours, the presentation of the stimuli and the scoring of the test. The original Stroop experiment found that it was harder to name patches of colour than it was to read written words, and that it was harder still to read the printed names of colours when the ink in which the named colour was printed was in a different colour (Lezak, 1995; Reynolds & Flagg, 1983) The finding led to a number of speculations about the meaning of this phenomenon. It was believed to relate to conflict in response demand, to the inability to inhibit an automatic response (reading), and to the inability to selectively focus on one stimulus whilst ignoring distractions (Lezak, 2004; Ponsford, 2000).

The Bohnen, Jolles and Twijnstra (1992) version of the Stroop Test was utilised in this study. The first two phases of the Stroop Test required the subject to read words and to name patches of colour at speed, respectively, and placed few demands on executive processes. The third phase of the Stroop test required subjects to read out a written colour name that was printed in incongruent ink at speed. This phase required the ability to focus attention by inhibiting an automatic response in favour of a novel response and to ignore distractions. The fourth phase of the Stroop test required subjects to switch between naming written colours and naming the printed colour of words at speed and subsequently placed more demands on executive processing. It measured the ability to focus attention, ignore distractions and rapidly switch attention between tasks (Lezak, 1995). Lezak (1995, 2004) noted that the addition of a mixed interference trial increased the complexity of the test. The mixed interference phase was not a part of the original Stroop task and involved more complex interference requiring the ability to switch attention (Bohnen et al., 1992).

The Trail Making and Stroop Tests were utilised to measure executive attention for this study because they were well-documented tests of attention that had been extensively validated in the literature (Lezak, 1995). They also did not include a memory component and could be labelled memory-free tests of complex executive attention. It was proposed that the way in which these tests related to tests of working memory would indicate which working memory tests tapped into complex executive attention, and which did not.

1.8.2 Computer-based assessment

The International Testing Commission (ITC) recently released international guidelines for computer-based and internet-delivered testing (ITC Guidelines, 2005). These guidelines have been adapted to shape the Health Professions Council of South Africa (HPCSA) policy on computer-based testing in South Africa. Key current topics of concern in the development of computer-based tests include technological issues, psychometric issues, evidence of equivalence between computer-based versions and paper-and-pencil versions of tests, human factors issues in the presentation of computer-based or internet-delivered material, and other administrative and test development concerns (HPCSA Guidelines, 2006).

Some of the advantages of computer-based testing include the fact that computerised tests have often been perceived to be more enjoyable than paper-based tests (Davies et al., 2005). This has often increased the level of motivation of the test-taker to perform on a test. Computer-based testing has also reduced the time that it takes to administer a test, which could help to diminish the effects of fatigue on test performance (Gur, Ragland, Moberg, Turner, Bilker, Kohler, Siegel, & Gur, 2001). Computerised administration has standardised the administration and instructions of a test so that data validity is improved (Davies et al., 2005) It has enhanced the precision of timing, which has decreased the variability in timing on test validity (Gur et al., 2001). It has increased inter-rater reliability as recording and scoring of tests has been more accurate. Computer-based scoring has also reduced data-handling errors and has therefore increased the security of test results, as test information has been better documented and protected. Computerised testing has also allowed for modification of certain tests to adapt to the level of knowledge or skill of the test-taker (Davies et al., 2005).

In spite of the fact that computerised administration has improved data validity, there have been some threats to the validity of computer-based tests. For example, the actual time that it takes between a subject's response and the recording of that response has occasionally undermined the accuracy of timing, one of the factors thought to increase reliability on computer-based tests (Gur et al., 2001.) The hardware and software used to run the test, including the size and clarity of the computer screen has sometimes produced variability in the subject's response (Davies et al., 2005). Computer-based scoring methods have occasionally been prone to error. Computerised testing has also at times caused uncertainty around the administrative requirements of a task as practise examples have not always adequately demonstrated test requirements (Gur et al., 2001). The type of response required by the subject on a computer-based test may have added unnecessary complexity to a task. The neglect of qualitative data and behavioural observation in computerised assessment has also reduced information about important factors that affect performance, such as lack of motivation or cooperation on a test (Bush et al., 2002).

Computer-based tests were utilised in this study for a number of reasons. First, the study aimed to explore the administration of, performance on, and potential threats to the validity of a set of computer-based cognitive-neuropsychological tests developed by a practising neuropsychologist in South Africa. A key reason for employing these tests was to begin to evaluate some of the issues around computer-based testing and test development in the South African context. Second, the *n*-Back Test was originally a visually displayed, computer-based test. It was therefore appropriate to compare it with other tests that were also computer-based. There were no alternate computer-based versions of these tests available in South Africa of which the researcher was aware. The tests used in the study were also developed to take cognisance of the disadvantages of computer-based testing within the South African context.

1.9 Rationale and aims

The relatively recent introduction of computer-based testing to the testing community has produced a unique set of considerations, particularly within the South African context. A key concern is that lack of exposure to information technology and varying levels of computer literacy could negatively affect performance for a large portion of the population on computer-based tests (Davies et al., 2005; Fisher, 2006). Even the ability to operate a computer mouse has been implicated as a factor affecting performance on speeded neuropsychological tests (Fisher, 2006). It was suggested that test anxiety, social, cultural and linguistic factors could also affect computer-based test performance (Davies et al., 2005). On the other hand, there was evidence to suggest that familiarising a subject with the test requirements through practise could lessen the impact of computer literacy on test performance (Davies et al., 2005)

Another important consideration in the South African context is whether there would be differences in performance on computer-based versions of traditional tests according to demographic variables such as gender, age, and level of education, or whether these tests remained relatively unaffected by such variables (Foxcroft & Roodt, 2005). Only a few studies have examined performance on neuropsychological tests and considered discrepancies in test performance attributable to gender, age, or level of education, and even fewer studies have investigated test-wiseness or language issues as factors that could prevent particular groups of people from functioning at an appropriate level in the testing context (Nadolne & Stringer, 2000; Anastasi & Urbina, 1997). This study aimed to investigate the effects of some of these variables on test performance and to determine the validity of the computer-administered tests (Lezak, 1995; Foxcroft & Roodt, 2005).

This study expected to find that variables such as test anxiety, computer experience, confidence using a computer and the speed with which one utilised a computer mouse would affect performance on the computer-based tests utilised in the study, particularly on speeded computer-based tests. The study also expected to find that demographic variables such as home language and level of education would affect performance on computer-based tests, as evidenced in prior research (Davies et al., 2005).

One of the primary concerns in the assessment of working memory was that research often focused selectively on either the storage or attentional facets of working memory tasks, and little attempt had been made to look at both these aspects together (Carpenter et al., 2000; Conway et al., 2005; Wager & Smith, 2003). This study aimed to investigate the relationship between storage of information and executive attentional processing within working memory. It also aimed to examine how tests of executive attention that did not include a memory component related to working memory tests that incorporated central executive processes. The memory tests utilised in the study each measured the ability to store and recall verbal, spatial or visual information. Tests of working memory span included an additional requirement in that the subject utilised domain-general complex executive attention by retaining information short-term whilst simultaneously switching attention to another task (Conway et al., 2005).

This study aimed to investigate whether there were stronger relationships within the tests according to domain-specific verbal, visual or spatial components of working memory, or between the tests according to executive attention. The tests administered in the study are well-recognised tests that have each been used extensively to measure short-term and working memory, and executive attention. However, the relationships between these tests have not been investigated previously in the literature. The study also aimed to evaluate whether working memory should be assessed using a number of tests to investigate different aspects of the construct, or whether only one test was necessary to measure working memory. A key aim of this study was to examine performance on the *n*-Back test, a relatively new test of working memory that had not been used in the South African context before, and to see how this test related to other tests of short-term memory span, working memory span, and executive attention.

Under the assumption that concurrent maintenance and manipulation of information was necessary for working memory capacity, and that this process utilised complex executive attention, this study expected to find that the tests used to measure executive attention, such as the interference and mixed interference phase of the Stroop Test and part B of the Trail Making Test, and tests measuring working memory capacity would meaningfully correlate. As the 1- 2- and 3-back conditions of the *n*-Back Test reportedly measured working memory capacity as processing load increased, it was hypothesised that the *n*- Back Test would reveal meaningful relationships with other tests of working memory capacity, such as the backwards condition of the Digit Span Test and the backwards phase of the Spatial Span Test (Logie, 1995; Rudkin et al., 2007). This study expected to find that the tests that measured short-term memory capacity would not have a significant relationship with measures of working memory capacity. Neither would they have a relationship with tests of complex executive attention such as the interference or mixed interference conditions of the Stroop Test, or with part B of the Trail Making Test. It was proposed that tests of short-term memory, such as the forwards conditions of the Digit Span and Spatial Span Tests would correlate according to the measurement of simple memory span.

A primary concern with neuropsychological testing was that it was often more difficult to detect mild to moderate cognitive impairment in adults who had high pre-morbid levels of education. These people tended to perform within normal limits even when impairment was apparent in everyday life. The cognitive reserve hypothesis maintained that level of pre-morbid cognitive ability partly determined the amount of cognitive loss, and the risk of developing dementia, post- injury (Howieson, Loring & Hannay, 2004). Education had been found to affect performance on almost all tests of cognitive ability, including tests that theoretically did not rely on education for performance. People with high pre-morbid levels of education tended to perform better on cognitive tests than people with lower levels of education (Howieson et al., 2004). In order to begin to address the issue of impairment in highly educated individuals, this research aimed to examine how healthy, well-educated, South African adults performed on tests of working memory.

This study aimed to examine the nature of the relationship between the domain-specific components of working memory outlined by Baddeley and Hitch, and the executive elements of attention in working memory. It aimed to investigate how working memory could be separated into different levels of storage and processing and how these levels operated together, and how performance on some well-known and well-used neuropsychological tests that tapped into short-term memory capacity, working memory capacity and executive attention compared to performance on the *n*-Back Test. It also explored the important and challenging question of computer-based cognitive-neuropsychological testing in South Africa.

1.10 Research Questions

These aims give rise to the following research questions:

1. What are the key issues around computer-based testing in the South African context?
2. Do tests of working memory reveal domain-specific relationships within each of the verbal, visual and spatial subcomponents of working memory, respectively, or do the tests relate according to domain-free executive processes?

3. Does the *n*- Back Test demonstrate adequate construct validity as a test of working memory?

1.11 Hypotheses

1. Computer mouse ability, computer experience, test anxiety, and confidence using a computer will affect performance particularly on speeded computer-based tests.
2. There will be some difference in performance according to gender, level of education, home language and confidence speaking, writing and reading in English on the computer-based tests.
3. The 1-Back, 2-Back and 3-Back conditions of the *n*- Back Test and the backward conditions of the Digit Span and Spatial Span Tests will correlate with part B of the Trail Making Test, and with the interference and mixed interference phases of the Stroop Test as tests that tap into domain-free complex executive attention.
4. The forward condition of the Digit Span Test and the forward condition of the Spatial Span Test will correlate with and factor together as measures of simple memory span.
5. The *n*-Back Test will correlate with and factor together with the backward condition of the Digit Span Test and the backward condition of the Spatial Span Test as measures of working memory span.

CHAPTER TWO

Methodology

2.1 Introduction

Neuropsychological tests have been used to operationalise a number of cognitive-neuropsychological concepts, such as working memory. Quantitative or mechanical methodology has focused on the statistical and psychometric interpretation of test results (Foxcroft & Roodt, 2005). This approach has been criticised for reducing human behaviour to a set of numbers. However, one of the key strengths is that statistical techniques have allowed for the detection of *patterns* of performance on a particular measure. This has allowed for conclusions to be drawn based not only on observation, but on the confirmation of observation in a set of consistent mathematical terms, leading to more accurate and generalisable results (Aron & Aron, 1999; Miles & Huberman, 1994).

Qualitative or non-mechanical methodology has focused on interpreting test scores in a holistic manner (Foxcroft & Roodt, 2005). Within this approach, demographic background, social environment, and test scores have been taken into account to put together a sense of the test-taker as a whole. One of the weaknesses of this approach is that there has tended to be some susceptibility towards exaggerating the importance of environmental influences and undermining test scores. There has been an element of subjectivity or relativity, which has demanded vigilance on the part of the assessor to be mindful of bias in interpreting test results. However, a key strength is that the analysis of test data could be concurrently interpreted with clinical information, leading to a more inclusive and holistic interpretation of test results (Foxcroft & Roodt, 2005).

To investigate the concept of working memory, inferential statistical methods were utilised to gain insight into the patterns of performance on a set of computerised cognitive-neuropsychological tests. This approach was employed for the purposes of standardising test administration so that performance within a group of subjects on the tests could be quantitatively compared and analysed. Qualitative methods were used in order to guide the theoretical development of the study and to guide the development of test instruments. The tests incorporated findings from previous research that the researcher conducted; a pilot study, which qualitatively examined the administration and procedure of each test. These findings were then communicated to the test developer and the tests were modified accordingly.

An unstructured pilot study was conducted with five volunteers to test the procedures for administration, scoring and computer-based presentation of stimuli. The pilot study investigated whether test instructions were easily understood, whether there were issues with the computer-based materials and to investigate floor and/or ceiling effects in the data (Harris, 1986). Floor and ceiling effects did appear in the data on some of the sub-components of the tests. However, as the researcher was testing the limits of working memory capacity, tests that produced floor effects were included. Ceiling effects were expected in some of the tests. The measures that produced these effects were often useful as a baseline of performance without cognitive load. Cognitive load was then added and the results could be compared to determine the relative effect of the cognitive demand. Test variables that did not have a significant relationship with any other variable, due to floor or ceiling effects, were identified so they could be removed from the statistical analysis.

The pilot study revealed that the written instructions on the *n*-Back Test were not suitable, as the subjects did not understand the requirements of the test. Subsequently, the written instructions were changed to facilitate better understanding of the test requirements and a practise trial was included to allow visual demonstration and practise on each phase of the test. The initial trials of the *n*-Back test included 200 stimuli per trial and this was reported to be too long. Pilot subjects lost motivation and in some cases refused to complete the test. The original *n*-Back design would have taken a total time of 47 minutes and 40 seconds to complete, excluding practise time. The test was subsequently altered to include 100 test stimuli, cutting the time to 23 minutes and 50 seconds to complete, excluding practise time. Problems with the computer timer included varying item presentation time and disruptions in the timing of speed-based tests. The test developer subsequently corrected these anomalies. The pilot study revealed the need to introduce practise examples on some of the tests so that the subject could familiarise him- or herself with the response requirements and mode of presentation of the tests. Practise examples were subsequently introduced on these tests.

2.2 Research Design

The study was separated into two parts. The first part investigated the effects of demographic variables and computer familiarity on test performance. The second part investigated the relationships between the tests and the validity of the *n*-Back Test. The research design was non-experimental as none of the variables were directly manipulated, and there was no control group and no random assignment. Direct causal relationships were accordingly not sought. Correlational techniques and Factor Analysis were used to investigate the relationships between the variables in the first part of the study. The second part of the study had an ex-post facto research design. The variables were artificially labelled as independent or dependent variables in this section. The study had a mixed design: it had a within subjects design in that all subjects received all test conditions, and a between-subjects design in that the results on each test condition were compared between the subjects of the sample. The study was descriptive, as the data obtained was examined and explained in light of theory. It was also cross-sectional as it was based on the observation of a number of variables occurring at the same point in time.

This study was primarily exploratory in nature. There were three key reasons for this. The first was to explore the computer-based administration of the tests used in the study so that the validity of the tests could be investigated. The second was to examine the effect of demographic variables on test performance as they naturally occurred in a sample. The third reason was that the study aimed to establish relationships between the variables, based on theory, which could then be used to investigate the construct of working memory more effectively in future research.

In the first part of the study, the independent variables were:

(1) Computer mouse ability;

(2) Computer experience (experienced/not experienced);
(3) Computer confidence (confident/not confident);
(4) Reported anxiety (anxious/not anxious);
(5) Age;
(6) Gender (male/female);
(7) Home language (first language English/second language English);
(8) Level of education (13-15 years/ 16-19 years);
(9) English confidence (confident/not confident);
(10) Order in which the tests were administered (Digit Span, Spatial Span, Trail Making, Stroop, *n*- Back/ *n*- Back, Stroop, Trail Making, Digit Span, Spatial Span)

The dependent variables in the first part of the study were:
(1) 0-back condition of the *n*-Back Test;
(2) 1-back condition of the *n*-Back Test;
(3) 2-back condition of the *n*-Back Test;
(4) 3-back condition of the *n*-Back Test;
(5) The word-reading phase of the Stroop Test;
(6) The colour-reading phase of the Stroop Test;
(7) The interference phase of the Stroop Test;
(8) The mixed interference phase of the Stroop Test;
(9) The forwards phase of the Digit Span Test;
(10) The backwards phase of the Digit Span Test;
(11) The forwards phase of the Spatial Span Test;
(12) The backwards phase of the Spatial Span Test;
(13) Part A of the Trail Making Test;
(14) Part B of the Trail Making Test;
(15) The computer mouse test.
These were also the variables in the second part of the study.

An Exploratory Factor Analysis (EFA) was conducted to investigate the factor structure underlying the variables in the study (Costello and Osbourne, 2005). Kaiser's Measure of Sampling Adequacy (MSA) was employed to determine which variables would be kept for the factor analysis. It was assumed that certain variables would not correlate with any other variable, as floor and ceiling effects were apparent in the pilot study. In other words, certain test variables were either too difficult or too easy and would consequently not reflect working memory ability. Another method employed to determine which variables would be kept for analysis was to examine the correlation matrix. This approach was based on the premise that if certain variables were uncorrelated with any other variable, and removal of the variable could be theoretically justified, the variable should be removed from further analysis.

2.3 Sample

The sample was a non-probability, convenience sample of 105 South African adults. The sample was initially meant to represent young adults. However, after the results had been checked for outliers, and the outliers had been removed, the sample still contained a few older subjects who performed within the normal range. The results of these older subjects were therefore included in the analysis. Seventy-nine subjects were female and twenty- six subjects were male. The subjects varied in age from 17 to 53, with a mean of 21.66 years (SD = 5.80). Seventy-five volunteers were recruited by the researcher from the first year psychology course at the university. Additional volunteers were recruited from second year, third year, Honours and Masters Psychology courses, and 'walk-in' students that had heard about the study and volunteered to participate. A total number of 109 volunteers participated in the study. However, 4 were excluded from the study as they reported a history of head injury with hospitalisation, learning disability and/or diagnosed neurological or psychiatric problems (McAllister, 2006a, 2006b).

Subjects had to have completed a *minimum* of 12 years of formal education (with matric exemption). The reason for this was twofold. First, one of the aims of the study was to explore the performance of educated (post-matric) adults on the neuropsychological tests administered in the study. Second, completion of grade 12 indicated that the subject had achieved a certain level of 'test-wiseness' that is reached within the formal education system. Davies et al. (2005) noted that test-wiseness is a function of the experience of formal schooling, whereby people learn particular thinking and reasoning strategies, problem solving skills, independent thinking, and the ability to work at pace and with accuracy. Test-wiseness has been described as “test-taking ability”, or the ability, motivation, skill and level of understanding of what needs to be done in order to perform effectively on a test (Davies et al., 2005).

The subjects were assumed to be appropriately fluent in English, as the completion of a matriculation certificate demanded a certain level of English proficiency. However, subjective level of comfort speaking and reading in English was recorded in order to ensure that this factor did not interfere with performance on the test battery. Only seven participants reported being *moderately* comfortable speaking, writing and reading in English, all other participants reported being *very* comfortable speaking, writing and reading in English, and *no participants* reported not being comfortable speaking, writing and reading in English. Second language English speakers included first language Afrikaans (6), Zulu (10), Tswana (2), Xhosa (2), Sotho (9), Shona (1), German (1), and Siswati (1). Years of education was divided into undergraduate (12 to 15 years) and postgraduate (16-19 years) and ranged from 12 to 19 years. Twenty-three students in the sample were postgraduate students whilst eighty-two were undergraduate students.

A questionnaire encompassing factors that could influence performance on the tests was administered in order to exercise statistical control over some of these potentially confounding variables. To improve the face validity of the study, a set of questions was put to each test-taker after completing the battery regarding the user-friendliness of the tests. The demographic characteristics for the sample are described in Table 1.

Table 1

Demographic Characteristics for the Sample

Demographic	Subdivision	Frequency	Mean/ Standard Deviation/ minimum-maximum
Order of tests	1 = Digit Span/Spatial Span/ Trail Making/Stroop / *n* -Back	44	-
	2 = N-back/ Stroop / Trail Making/Digit Span/ Spatial Span	61	-
Gender	Male	26	-
	Female	79	-
Home Language	1st language English	73	-
	2nd language English	32	-
Level of Education	12 –15 years	82	12 – 19 years
	16-19 years	23	
Age	-	-	*M* (21.66) *S* (5.80) 17- 53 years
Reported Anxiety	Anxious	38	-
	Not anxious	67	-
Experience using a computer	Experienced	82	-
	Not experienced	23	-
Confidence using a computer	Confident	84	-
	Not confident	21	-
Confidence speaking, reading and writing in English	Very confident	95	-
	Moderately confident	10	-

a. Materials and Apparatus

i. Questionnaire

A computer-based questionnaire (Appendix A) was administered prior to testing. The questionnaire was developed by the author following two years supervised training in administering neuropsychological tests, and by a practising clinical neuropsychologist in the South African context, who developed the computer-based tests that were utilised in the study. The questionnaire comprised 13 items. Each question was presented singly in the centre of a laptop computer screen placed directly in front of the subject. The subject used a computer mouse to select an answer from a number of prescribed answers presented below each question. The response was recorded via the computer and there was no time limit for viewing or answering the questions.

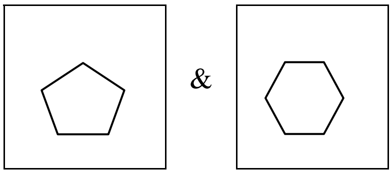
ii. Computer Mouse test

A computer-based test of computer mouse ability in which the subject's ability to click a computer mouse on a series of targets in order, to click and drag various shapes on a computer screen, and the speed at which a subject could do so, was administered following the questionnaire (Appendix B). The mouse test was presented on a laptop computer screen placed directly in front of the subject and comprised two subtests. Part A involved clicking on the numbers 1 to 20 in order, which were presented in a block formation in the centre of the computer screen. Part B involved dragging and dropping 20 items presented in the same block formation in the centre of the screen that resembled blocks of cheese onto a picture of a mouse presented at the bottom of the screen. The spatial layout of the items was identical in part A and part B so that differences in time taken to complete each part could not be attributed to differences in distance between the items. The time taken to complete each subtest in seconds and milliseconds was recorded via the computer and the sum of the total time taken to complete both subtests was used as a measure of computer mouse ability in the study. Error was accounted for in that the subject could not move onto the next item in the test until the current item was correctly responded to. In this way, error was assimilated into the amount of time that the subject took to complete the test.

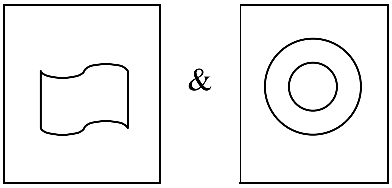
iii. *n*-Back Test

The *n*-Back Test was a sequential memory task in which the subject was required to examine a continuous stream of items and report whenever an item corresponded to one that was presented *n* items previously in the stream (Conway et al., 2005). The present study used a visual computer-based version of the *n*-Back Test (Appendix B). This version of the test had four different conditions, 0-back, 1-back, 2-back and 3-back (Perlstein et al., 2001). In this version, a series of items was presented at a rate of one per 1500 milliseconds in the centre of a laptop computer screen. The inter-stimulus interval ranged from 1000, 1500, 2000, to 2500 milliseconds across the conditions, respectively. The centre stimulus presentation block remained blank during the inter-stimulus interval. In the 0-back condition, subjects were instructed to press the space bar on the laptop keyboard in response to a particular predetermined target. The target in this condition was a yellow cross. The 1-back condition involved pressing the space bar in response to any item that was identical to the one directly preceding it. The 2-back condition involved pressing the space bar to any item that was identical to the one that was presented two items back. The 3-back condition involved pressing the space bar to any item that was identical to the one that was presented three items back.

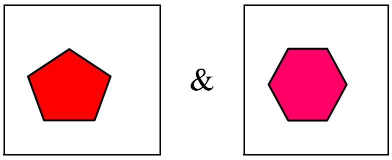
The *n*-Back Test stimuli were 10 unique shapes in unique colours and they were presented in the same random sequence for each subject on each trial. Performance was documented as the number of correct responses out of 20 possible matches (accuracy), although the number of omission and commission error was also recorded. The *n*-Back Test designed for the study attempted to prevent recoding of visual to verbal information by presenting shapes that were difficult to verbalise, particularly within the time limit imposed by the test. Some of the shapes and colours were recognisable and were therefore able to be recoded into a verbal format (e.g. square/circle/triangle and red/yellow/pink). However, a number of steps were taken to prevent the verbal recoding of visual stimuli. First, shapes were included that were very similar to other shapes in the sequence, making it difficult to distinguish between them under a time constraint on the basis of shape. For example,



Second, unusual shapes were included that were difficult to verbalise quickly due to the time restriction. For example,



Third, colours were included that were similar to the colour of other items in the sequence, making it difficult to distinguish verbally on the basis of colour. For example,



Test Characteristics

The *n*-Back Test has been used extensively in neuroimaging research. However, test characteristics for the *n*-Back Test were not readily available in the literature (Kane et al., 2007). Age was shown to affect performance in a study comparing performance on the *n*-Back Test in a sample of 68-year-olds and a sample of 20-year-olds, and in a study comparing performance in a sample of 70-year-olds and a sample of 30-year-olds. Age differences were apparent on all conditions of the test (Lezak, 2004). In another study, mild Traumatic Brain Injury (TBI) patients did not perform differently from control subjects. However, fMRI revealed that TBI patients used more extensive areas of the brain during performance on the *n*- Back Test, which suggested that they may have used more cognitive resources than control subjects to perform the task (Lezak, 2004).

One of the aims of this research was to examine the validity of the *n*-Back Test as it related to other measures of working memory, thereby adding to the literature.

iv. Digit Span Test

The present study used a visual computer-based version of the Digit Span Test (Appendix B). The Digit Span Test included in the Wechsler Adult Intelligence Scale (WAIS-III) batteries was made up of two tasks, each of which involved different mental processes (Lezak, 1995, 2004). On digits forward, an examiner read aloud a series of random numbers, at a rate of one per second, increasing in length. The subject was required to repeat the digits, in the same order, immediately after the examiner had read them (Groth-Marnat et al., 2000). If the subject repeated the sequence correctly, he or she was then given the next sequence of numbers, increasing by one number. This carried on until the subject failed to repeat two sequences of the same amount of numbers in a row, or until the subject repeated a sequence of nine numbers accurately (Lezak, 1995, 2004). On the backward phase of the test, a fixed random series of numbers was read at a rate of one per second and the subject was required to repeat the series backwards, i.e. from the last digit to the first in the correct order (Groth-Marnat et al., 2000). The same number of items in a sequence, the same order of sequences and the same rules applied as on the forward phase of the test. The average difference between forward and backward scores was usually no more than two points in neurologically intact subjects with a three-point or more discrepancy occurring more often in subjects with brain damage (Lezak, 1995).

The version of the Digit Span Test used in this study was based on the WAIS - III version but differed slightly. First, the test was computer–based and the numbers were automatically verbalised via soundcard on a laptop computer placed directly in front of the subject at a rate of one per second. The test produced different sequences of random numbers for each trial (none of the numbers was repeated in a given sequence) for each subject. Responses were recorded as the physical entry of the digits into the computer keyboard. The forward phase started with a sequence of five digits whilst the backward phase started with a sequence of three digits.

Test characteristics

Scoring of the WAIS III Digit Span Test involved summing scores from the forward and backward conditions to arrive at one raw score. However, the tendency to combine scores on the two parts of the Digit Span Test for statistical analysis was flawed as it threatened the loss of important clinical information. It was consequently not sensible to convert raw scores on the Digit Span Test into standard scores (Groth-Marnat et al., 2000; Lezak, 1995, 2004). It was determined that performance on the forward and backward phases should be measured independently to investigate the different processes engaged by each part of the test (Lezak, 2004)

The first part of the test, Digits Forward, was a test of immediate serial recall, short-term memory storage capacity, and the ability to sustain the focus of attention (Baddeley, 1998b; Groth-Marnat et al., 2000). The normal range for digits forward was 6 plus or minus one (Lezak, 1995, 2004). Lezak (1995, 2004) noted that scores of six and above were comfortably within normal limits, that a score of five was within marginal to normal limits, that a score of four was borderline, and that a score of three indicated dysfunction. The second part of the test, Digits Backward, was a test of concurrent maintenance and manipulation of information, and working memory capacity (Groth-Marnat et al., 2000; Lezak, 2004). Scores of four or five on the backward phase of the test were within normal limits, scores of three were borderline substandard or substandard, depending on the subjects' level of education whilst a score of two indicated dysfunction in any subject group. The difference in raw scores between Digits Forward and Digits Backward ranged between 0.59 and 2.00 for neurologically intact subjects (Lezak, 1995). A difference of five or more digits between the forwards and backwards phase of the test was considered abnormal (Groth-Marnat et al., 2000)

Level of education affected performance on this task (Lezak, 1995, 2004). Age was associated with Digit Span performance so that, whilst forward span tended to remain steady, backward span tended to decline over time. However, it affected performance modestly and only past the age of 65 or 70 (Lezak, 1995). Performance on the forward phase of the test was only modestly affected after the age of 65 (Lezak, 2004). Anxiety also played a role in performance on the Digit Span Test in that it tended to decrease the amount of digits that the subject was able to recall (Lezak, 1995, 2004). Practise has been shown to decrease anxiety and improve performance on digits forwards (Lezak, 2004). Practise effects on the Digit Span Test have been noted as being moderately but statistically significant ($p < .045$), with test-retest reliability coefficients varying between .66 and .89 (Lezak, 1995). The Digit Span subtest of the WAIS III – Revised (WAIS III - R) was highly correlated with the WAIS-III-R working memory index (.83). Internal consistency coefficients for the working memory index averaged .94 (Groth-Marnat et al., 2000). For the WAIS III factor structure, the Digit Span Test loaded on working memory at .71 (Groth-Marnat et al., 2000).

Comparison between a computer-based version and the WAIS-III version of the Digit Span Test revealed convergent validity correlations of .53 between the computerised and original version of the test (Paul et al., 2005). Level of education has been found to affect performance on a computerised version of the Digit Span Test, with increasing levels of education being associated with higher test scores (Paul et al., 2005).

v. Spatial Span Test

A computerised version of the WAIS-III Spatial Span Test was utilised for the study (Appendix B). The Spatial Span Test was a variant of Corsi's Block Tapping Test, in which the subject was presented with nine 1 and a half- inch cubes arranged in random order on a flat surface. The experimenter tapped the blocks in a particular sequence and the subject was required to tap the blocks in the same sequence once the experimenter had finished. The layout of the cubes resulted in sequences that varied in length and in spatial organization (Lezak, 2004). The test was similar to the Digit Span Test in that the subject's performance was given by the longest sequence that was correctly reproduced (Baddeley, 1996c). In the computer-based version, the subject was presented with nine two-dimensional squares placed in a fixed random position in the centre of a laptop computer screen placed directly in front of the subject. Responses were recorded as the physical clicking of a computer mouse on the blocks in the correct order to that in which they lit up on the computer screen. On the backward phase of the test, responses were recorded as the physical clicking of a computer mouse on the blocks in the reverse order to that in which they lit up. In other words, subjects had to click on the last block and work backwards to the first. The spatial layout of the items was identical on the forward and backward phases of the test.

Test Characteristics

Test characteristics for Corsi's Block Tapping Test are given, as it was methodologically and theoretically analogous with the Spatial Span Test administered in the study. There was a difference of at least one item between Digit Span and Corsi Span in neurologically intact subjects (Lezak, 1995, 2004). However, it has been reported that the Digit Span Test was usually about two items higher than Corsi Span in healthy control subjects (Lezak, 2004). Baddeley (1996c, 2003) reiterated the finding that performance on the Digit Span was usually about two items more than the Corsi Span. Lezak (1995) noted that the spatial layout of the Corsi Test influenced performance, with sequences that contained equal distances more liable to be recalled accurately and sequences with shorter distances more liable not to be recalled accurately. Education has been found to influence performance, as discovered in a study in which one third of the subjects were educated below sixth grade (Lezak, 1995, 2004). Gender affected performance to a small degree, with men scoring moderately (around .30 of a point) but significantly ($p < .001$) higher than women, but not when level of education exceeded 12 years. Age did not affect performance until after 60 years, at which point performance progressively declined (Lezak, 1995). Comparison between a computerised version and a standard neuropsychological measure of the Spatial Span Test (WAIS-III) revealed convergent validity correlations of .63 between the tests (Paul et al., 2005).

vi. Trail Making Test

A computer-based version of the Trail Making Test (TMT) was utilised for the study (Appendix B). The computer-based version of the TMT presented a randomly distributed set of numbers printed in lower case in the centre of a laptop computer screen placed directly in front of the subject. The spatial layout of the items was identical on part A and part B so that differences in time taken to complete each phase could not be attributed to differences in the distance between the items. Responses were recorded as the clicking of a computer mouse on the numbers 1 to 25 in order on part A of the test and the clicking of a computer mouse alternating between numbers and letters of the alphabet in the correct order, respectively, on part B. Errors were recorded but were only important insofar as they contributed to a slower overall response time (Broshek & Barth, 2000). On the computer-based Trail Making Test the subject could not move on to the next item until the correct response was given on a current item. Time taken in seconds and milliseconds to complete each part of the test was used as a measure of performance.

Test characteristics

The TMT is part of the Halstead-Reitan Neuropsychological Test Battery. The psychometric properties of this battery have been explored more comprehensively than any other neuropsychological battery (Broshek & Barth, 2000). The TMT is reliant on complex visual scanning, speed and coordination in terms of motor function. The test has a strong motor constituent in that speed of motor performance and dexterity contribute convincingly to success on this test (Lezak, 2004). It is particularly sensitive to the effects of brain injury, as are other tests that involve attention and speed of motor functioning (Lezak, 1995). Reliability coefficients for the TMT generally varied between .60 and .90. Low reliability coefficients were often found on Part A with schizophrenic patients (.36), while exceptionally high scores on Part A have been found in patients with vascular problems (.94) (Lezak, 1995, 2004). Practice effects have been found on the TMT on retesting. However, only improvement on Part A tended to be statistically significant as group variances on Part B were typically considerable (Lezak, 1995, 2004). When re-administration of the test occurred one year later, no practice effects were found (Lezak, 2004). When re-administrations of the test were placed one week to three months apart, part B revealed significant practice effects, but not when the test was administered three months later (Lezak, 2004). Age has played a role in TMT performance; normative research has demonstrated that time taken to complete the TMT increases with each decade (Lezak, 1995, 2004).

Scores on Part B of the TMT have been strongly correlated with level of education and intelligence (Broshek & Barth, 2000; Lezak, 2004). Gender did not seem to affect performance significantly on this test. However, there has been some evidence to suggest that women, particularly elder women, generally perform slower than men on Part B of the test (Lezak, 1995, 2004). There are strong links between Part B of the TMT and both attention and conceptual reasoning measures and a strong link between Part B and visuospatial intelligence (Groth-Marnat, 2000). Depression had an effect on Part B of the TMT, in that performance was abnormally slowed. Both parts of the TMT were sensitive to the effects of cognitive impairment in dementia. Performance on the TMT by patients with traumatic brain injury was slower than that of control subjects, and slowing increased with the level of brain trauma incurred (Lezak, 1995). Comparison between a computerised version and a standard neuropsychological measure of the TMT revealed convergent validity correlations of .53 between part A, and correlations of .65 between part B of the computerised version and a paper-and-pencil version of the Trail Making Test (Paul et al., 2005).

vii. Stroop Colour Word Test

A version of the Stroop Colour Word Test, developed by Bohnen, Jolles and Twijnstra (1992), was adapted for computer use in the study (Appendix B). The Bohnen et al. (1992) version of the Stroop Test had four subtasks. The first was naming of colours written as words in black ink (Stroop1). The second task was naming of colours presented in patches of coloured ink (Stroop2). The third task was an interference task in which the subject had to name the colour of the ink in which a word was printed, ignoring the fact that the written word was incongruent (Stroop3). The fourth task was a mixed interference task in which the subject had to switch from naming the colour of the ink in which a word was printed to naming the written word (Stroop4). This was accomplished by placing a black rectangle around twenty random words. If the word did not have a black rectangle around it, the subject was required to name the colour that the word was printed in and ignore the written word. If a rectangle appeared around the word, the subject had to read the word within the rectangle as opposed to naming the colour of the ink in which the word was printed (Lezak, 1995, 2004).

The computer-based Stroop Test in the study presented each item one at a time in the middle of a laptop computer screen placed directly in front of the subject. There were 100 items on each phase of the test (Lezak, 1995). Four colours were used– red, blue, yellow and green. The computer-based version of the Stroop Test measured performance in terms of time taken to complete each phase of the test, individually. The test used in the study did measure error separately, but error factored into the time taken to complete each phase of the test. Time was consequently used as a measure of Stroop performance.

Test Characteristics

With regards to the psychometric properties of the Stroop task, reliabilities ranging between .73 and .85 have been reported on the individual version of the task with reliability on the interference score at .70 (Groth-Marnat, 2000). These figures indicated strong reliability on a version of the Stroop Test. Raw scores revealed good test-retest reliability while interference effect reliability was average. However, practise effects have been demonstrated many times on the Stroop Test (Groth-Marnat, 2000). Intelligence has been correlated with the Stroop test and may have an influence on performance. Faster completion times were positively correlated with higher intelligence (Groth-Marnat, 2000). Reading proficiency was demonstrated to be integral to the Stroop interference effect and therefore subjects were required to have single-word reading capacity in English (Groth-Marnat, 2000). Anxiety affected performance on the Stroop Test, affecting men more than women (Lezak, 2004). However, in the absence of anxiety, men performed faster than women on the word-reading, colour-naming and interference trials. Significant differences in performance between males and females were not found in a large study (Lezak, 2004). Slower performance was found with increasing age across a number of studies (Lezak, 2004).

Comparison between a computerised version and a standard neuropsychological measure of the Stroop Test revealed convergent validity correlations of .70 between the interference trials of the computerised and standard version of the test (Paul et al., 2005).

b. Procedure

i. Procedure for the Study

Having gained ethical clearance from the university (ethics clearance protocol number 70205), and permission from the registrar to approach students for the purposes of the study, the researcher subsequently obtained written clearance from the head of school of the department of psychology to approach students within the department. Students were approached in their lectures by the researcher, and the purpose and exclusion criteria of the research were explained verbally and in an information sheet (Appendix C). Those who were willing to volunteer wrote their names on an appointment list for an appointment at a time that suited both the researcher and the participant. Once a subject had agreed to participate, and had arrived at the venue, the research was described in more detail before an informed consent form was signed and testing commenced.

The researcher underwent two years of informal training in neuropsychometric testing and assessment at a rehabilitation hospital under the supervision of a registered clinical neuropsychologist. She worked extensively with both paper-and-pencil and computer-based tests. The importance of environmental factors was acknowledged and testing conditions were standardised as much as possible. Subjects were requested to turn off their cell phones and to put on corrective glasses if necessary before testing commenced. The researcher administered each computer-based test individually to each subject in a quiet room that was assigned to for testing in the psychology department, or in a quiet room with adequate lighting at a venue that was convenient to the subject. All the tests were administered in English, as it was assumed that subjects had either passed English sufficiently in school, or had passed a university English proficiency examination prior to acceptance.

The computer-based tests, listed above, were all administered using a laptop computer with a 30cm screen. Three laptop computers were used. Two were Acer Pentium 4 laptops, and one was a Hewlett Packard laptop. The tests were designed to run on Microsoft® Windows® 2000. The tests were administered to each participant one at a time on full volume. For all of the subjects, the Questionnaire and Computer Mouse Test were administered first in that order. For some of the subjects, the Digit Span Test was administered next, followed by the Spatial Span Test, the Trail Making Test, the Stroop Test and the *n*- Back Test respectively. For the remaining subjects, the tests were administered in an opposite order. The tests were reordered to ensure that fatigue was not a factor during testing. The researcher drew lots to determine which order the subject would receive. All of the computer-based tests administered in the study had standardised auditory or visual instructions and practice tasks so the subjects could familiarise themselves with the requirements of the test and with the mode of response for each task.

ii. Procedure for each test

Computer Mouse Test

The computer mouse test was administered first. This test had a visual display and test instructions were pre-recorded and played back via soundcard. The subject could not physically practise the test as the researcher wanted to gather a measure of the subject's baseline ability to use a computer mouse (Davies et al., 2005). The instructions for the mouse test were displayed in a speech bubble in the top left hand corner of the screen. The instructions for part A required the subject to click via computer mouse one at a time on a set of numbers presented in a set of blocks from 1 to 20 in order. The subject was instructed to complete the task as quickly and accurately as possible and was allowed a maximum of three attempts on each phase of the test. The subject was allowed a maximum of ten errors on each phase after which the phase was aborted.

The instructions for part B of the test required the subject to drag and drop a set of blocks one at a time onto the picture of a mouse presented at the bottom of the screen in any order. The subject was instructed to complete the task as quickly and accurately as possible and was allowed a maximum of three attempts on each phase of the test. The subject was allowed a maximum of ten errors on each phase after which the phase was aborted. The subject had to complete each phase of the test 100 percent correctly to move onto the next test. All of the subjects completed the test in fewer than five minutes and with less than ten errors. The computer mouse test was scored as the sum time taken to complete part A and part B, in seconds and milliseconds.

n-Back Test

The *n*-Back Test included 20 practise stimuli and 100 test stimuli on each phase of the test. Subjects could attempt each practise trial a maximum of four times, after which the phase was aborted. The subject had to complete the practise trial 100 percent correctly on the 0-Back and 1-Back conditions, and 80 percent correctly on the 2-Back and 3-Back conditions to participate in the corresponding phase. Visually displayed instructions were presented in a speech bubble in the centre of the computer screen, to the left of the visual display of the test stimuli. Subjects were instructed to press the space bar whenever they saw an item that matched one that was presented *n* items back, according to the instruction for that trial. Subjects were instructed that when a response was correct, the button featured below the stimulus would flash green, indicating a correct response, and when a response was incorrect it would flash red, indicating an incorrect response, in order to give visual feedback to the subject. When there was no response on the part of the subject, the button below the stimulus remained blank. During the inter-stimulus interval, the stimulus button also remained blank. The practise trials gave additional visually displayed written feedback when the subject made a correct or incorrect response to facilitate understanding of the requirements of the test. They were either informed “Yes, that is correct” or “No, that is incorrect”, followed by repetition of the original test instructions.

On the 0-Back phase of the test, subjects were instructed to press a spacebar in response to a yellow cross whenever they saw it in a sequence of stimuli. They were instructed by the examiner not to respond to any other item in the sequence. On the 1-

Back phase of the test, subjects were instructed to press the space bar when they saw an item that was identical to one presented 1 item ago. On the 2-Back phase of the test, subjects were instructed to press the space bar when they saw an item that was identical to one presented 2 items ago. On the 3-Back phase of the test, subjects were instructed to press the space bar when they saw an item that was identical to one presented 3 items ago. Further information, given by the examiner, informed subjects that they should only respond to an item that was identical in both shape and colour according to instructions for that phase of the test. On each phase of the test, subjects were instructed by the examiner not to respond to any other item in the sequence that did not match the one presented *n* items back. The test was scored as the number of items correctly identified out of 20 target stimuli on each phase of the test.

Digit Span Test

The Digit Span Test in the study included a practice trial before the start of the forward phase of the test. Test instructions were visually displayed in the centre of the computer screen to the left of the displayed stimuli. On the practice trial, the subject had to listen to the numbers “1, 2, 3” read out by the computer via soundcard at a rate of one per second, and then punch in the numbers 1-2-3 on a laptop keyboard in the correct order. Subjects were instructed to wait until the auditory presentation of the stimuli stopped and the response screen was presented before entering the numbers. The computer did not register the numbers until the response screen was presented after all the numbers in the sequence had been read out. Subjects were instructed that they were able to correct a mistake in the sequence by pressing the backspace button on the keyboard once the digit had been entered, to allow for the correction of errors. To correct a digit in the sequence, subjects were instructed to click on the block that held that number, press the backspace key, and fill in the correct number. The subject could visually view their response before clicking the ‘next’ button on the computer screen. The practice phase allowed a maximum of three trials in which the subject had to complete the practice trial 100 percent correctly, after which the phase was aborted.

Having successfully completed the practice trial, the forward phase of the Digit Span Test was administered. The forward phase began with 5 digits rather than three, based on Lezak's (1995) assertion that it was not necessary to start with three digits with neurologically intact subjects. If the subject could correctly enter the digits in the correct order within two attempts, the test moved on to administer 6 digits. The subject had a maximum of two attempts on each sequence before the test was aborted. If the subject could not enter 5 digits in the correct order on the first sequence, the test reverted to a sequence of 4 digits. If four digits were not entered in the correct order, the test reverted to a sequence of 3 digits. The subject was instructed to punch in the numbers on the keyboard in the order in which they were given. The subject also had the freedom to change their response if he or she did not feel that the initial response was correct. When the subject was satisfied that the sequence was correct then they could press the 'next' button, which moved the subject on to the next sequence.

The backward phase of the Digit Span test administered visually presented instructions but the test developer did not include a practice trial on this phase of the test. Written instructions informed the subject that they had to reverse the order of a sequence of digits, starting at the last digit and working backwards towards the first. The subject was given an example that if the sequence "1, 2, 3" was presented, the correct response would be "3, 2, 1". The sequence began with the computer-articulated verbal presentation of three digits, and the participant was instructed to reverse the order of the digits and punch in the numbers on the keyboard from the last to the first number in the sequence. The participant was observed closely by the examiner to ensure that he or she did not enter the numbers in a 'forward' sequence starting from the right and moving left. In every other manner, the backward phase was administered in the same way as digits forward. The test was scored as the maximum amount of numbers given in the correct sequence on the forward and backward phases of the test.

Spatial Span Test

The Spatial Span Test administered visual instructions that were presented in the centre of the screen to the left of the displayed stimuli. The subject first completed a practice trial for the forward phase of the test. A fixed set of nine two-dimensional 'blocks' was presented on a computer screen. Three blocks then 'lit up' in a particular order, indicating spatial movement. The subject was instructed to click on the blocks in the same order in which they lit up, from first to last. The subject was instructed prior to the test to wait for the sequence of stimuli to stop and for a response screen to be presented before they could click on the blocks in the order in which they lit up. The computer did not register responses that were made by the subject before the response screen was presented. Numbers appeared in each block in the order in which the subject responded. Number cues had been found to track performance on the Spatial Span Test more easily (Lezak, 2004). Subjects were instructed that they could correct an incorrect spatial sequence by double-clicking on the block that they had incorrectly clicked on, which then removed the number from the block, to allow for the correction of errors. The subject had a maximum of three practice trials, in which the correct sequence had to be replicated in order, after which the phase would be aborted.

Having successfully completed the practice trial, the forward condition of the Spatial Span test was administered. The test began with five grey-coloured blocks that each 'lit up' in a blue colour one at a time at a rate of one per second. As the participant responded by clicking on each block, a number appeared in block in which the subject had clicked to allow the participant to view the order in which they clicked on the blocks. If the participant could correctly click on the blocks in the correct order within two attempts, the test moved on to administer a sequence of six blocks. The participant had a maximum of two attempts on each sequence before the test was aborted. If the participant could not enter five blocks in the correct order on the first sequence, the test reverted to a sequence of four blocks. If four blocks were not entered in the correct order, the test reverted to a sequence of three blocks. The test was scored as the maximum number of blocks given in the correct sequence. The backwards phase of the Spatial Span test began with the visual presentation of three blocks that lit up in a particular sequence, and the participant was instructed to reverse the order of the sequence and click on the blocks from the last to the first block that lit up in the sequence. The subject first completed a practice trial for this phase. The subject had a maximum of three practice trials, in which the correct sequence had to be replicated in the reverse order, after which the phase would be aborted. In every other manner, the backward phase was administered in the same way as spatial forward.

Trail Making Test

The TMT was given in two parts; part A and part B (Lezak, 1995). In the most commonly used version of the test, part A contained circles enclosing the numbers one to twenty-five scattered randomly across a page. Participants were required to link these numbers in order by drawing a pencil line from number to number as quickly as possible. Time to completion was recorded in seconds. If errors were made, the assessor quickly pointed them out and redirected the test-taker to the next correct number. Part B contained circles enclosing both numbers and letters of the alphabet. The test-taker was required to alternate between numbers and letters of the alphabet consecutively, linking the first number with the first letter, the first letter with the second number, the second number with the second letter, and so forth. Again, if errors were made the assessor pointed them out and redirected the test-taker to the next correct number or letter in the sequence.

The TMT in this study administered a practice trial prior to each part of the test. Test instructions were visually displayed in the centre of the screen to the left of the test stimuli. On the practice trial for part A of the test, the participant was instructed to click as fast as possible on the numbers 1 to 10, moving in numerical order. Once the subject had clicked on the number 1, a red line with an arrowhead appeared which the participant then connected to the next number. If the participant clicked on the next correct number in the sequence, the line turned blue. If the participant clicked on an incorrect number in the sequence, the line would not join on to that number, and did not turn blue. Error was noted via an unobtrusive sound that indicated that the incorrect item had been selected. The error sound persisted each time the subject selected an incorrect item until the subject had selected the next correct item in the sequence. The test could only progress once the participant had corrected their error, and the line was joined to the correct number. Each time that the subject clicked on an incorrect number, written instructions appeared on the screen that reported, “No, that is incorrect" and then repeated the test instructions. The subject had to get the sequence 100 percent correct on the practice round, after which they moved onto the actual test.

The instructions for part A of the test required the subject to click on the numbers 1 to 25 in numerical order. The subject was instructed to complete the test as quickly and as accurately as they could. The instructions for part B of the test required the subject to click on numbers and letters in the following sequence: 1-A-2-B-3-C. The differences in spatial arrangement of the numbers and letters on part B of the test had been shown to make part B of the test more difficult than part A (Lezak, 2004). On the computer-based version of the test administered in this study, the numbers and letters were presented in exactly the same position on the screen as part A of the test. This made the distance between the stimuli identical so that differences in test performance could not be attributable to differences in spatial positioning of the items. Time was recorded in milliseconds and error factored into time, as the subject could not move on to the next item until the current one was correctly responded towards. The subject had to respond to every item in the test. The test was scored as the amount of time taken to complete part A and part B, individually, in seconds and milliseconds. The method of using time scores as a measure of performance in the place of standard scores has been put forward as a more suitable measure of performance on this task (Lezak, 2004).

Stroop Test

In the computer-based Stroop Test, test instructions were visually presented in the centre of the screen to the left of the displayed test stimuli. Each subject completed a practise trial of 10 items before each phase of the test. Subjects had to complete the practise trial 100 percent correctly, after which they moved onto the test. On the practise phase of each test, subjects were instructed to click on one of four "buttons" that matched a written word or ink colour printed above the row of buttons, presented in the middle of a computer screen. In the practise round and during the test, if the participant made a correct response a smiley face,☺, was presented and the computer moved on to the next stimulus. If the participant was incorrect a sad face,☹, was presented and the subject could not move on to the next stimulus until the correct response was made. The “faces” were displayed for approximately half a second between each stimulus.

The method of presentation of the stimuli was perhaps the most changed aspect of the computer-based version of the test used in the study. The stimuli appeared in the centre of a computer screen against a grey background. Below the stimulus was a set of four boxes, each containing the name of a colour printed as a word in black against the background of its corresponding colour, for example:

BLUE **GREEN** **RED** **YELLOW**

The subject was required to click via computer mouse on whichever box contained the correct response on the screen– either the name or the colour of the ink - according to instructions given before the test. On the word-reading phase subjects were instructed to click on the box that contained the word that matched the word presented in the centre of the screen. On the colour-naming phase subjects were instructed to click on the box that contained the colour that matched the patch of colour presented in the centre of the screen. On the interference phase subjects were instructed to click on the box that contained the colour that matched the colour of the ink in which a word was presented in the centre of the screen. However, this time the word was incongruent to the colour in which it was presented for example, **RED**. On the mixed interference phase the subject was instructed to click on the box that contained the colour that matched the colour of the ink in which a word was presented in the centre of the screen. However, if a black rectangle was placed around the word, the subject had to read the word and to ignore the colour of the ink in which the word was printed, for example,

RED

On the computer-based Stroop Test the subject could not move on to the next item until the correct response was made on a current item. The subject had to respond to every item in the test. The test was scored as the amount of time taken to complete each phase of the test, individually, in seconds and milliseconds.

c. Ethical Considerations

Ethical clearance was obtained from the university ethics committee for postgraduate, non-medical research. Once a subject had arrived at the testing venue, written informed consent was obtained from each volunteer, following ethical guidelines. Subjects were informed that participation in the study would be of no direct benefit to the subject. Subjects were informed that there were no legal, psychological, financial or physical risks to participating in the study. Confidentiality was assured in that no person other than the researcher would have direct access to the subject's identifying details, and that all information that could identify the subject would be removed from the research report. Subjects were informed that their identifying data would be destroyed once the project had been completed and the project was marked. Subjects were informed that the research results would be reported in a Masters dissertation, and that a summary of the results of the sample would be available on the testing room door once the project was completed. Subjects were informed that group results could also be reported in a journal article.

CHAPTER THREE

Results

3.1 Introduction

This study explored the construct of working memory, utilising a selection of tests to operationalise the construct. Four computer-based cognitive-neuropsychological tests were administered to each subject in the sample to investigate five hypotheses. The first hypothesis was that computer mouse ability, confidence using a computer, experience using a computer and test anxiety would affect performance particularly on speeded computer-based tests. The second hypothesis was that there would be differences in performance according to level of education, home language and confidence speaking, writing and reading in English on the computer-based tests. The third hypothesis was that the 1- 2- and 3-back conditions of the *n*-Back Test would correlate with performance on the backward phase of the Digit Span Test, the backward phase of the Spatial Span Test, part B of the Trail Making Test, and the mixed interference phase of the Stroop Test as measures of complex executive attention. The fourth hypothesis was that performance on the forward phase of the Digit Span Test, and the forward phase of the Spatial Span Test would correlate and factor together as measures of simple short-term memory span. The final hypothesis was that the *n* -Back Test would correlate and factor with performance on the backward condition of the Digit Span Test and the backward condition of the Spatial Span Test as measures of working memory span.

The test of computer mouse ability included a clicking phase and a dragging and dropping phase. Performance on the computer mouse test was measured as the total time taken to complete both phases in seconds and milliseconds. The Stroop Test included four phases, each of which was measured individually in terms of time taken to complete each phase in seconds and milliseconds. The Digit Span Test included a forward and a backward phase. Each phase was measured individually in terms of the maximum number of digits entered in a correct sequence. The Spatial Span Test included a forwards and a backwards phase. Each phase was measured individually in terms of the maximum number of spatial items entered in the correct sequence. The Trail Making Test was given in two parts; part A and part B. Performance on each part was measured individually in terms of time taken to complete each phase in seconds and milliseconds. The *n*-Back Test was given in four phases and performance on each phase was measured in terms of number of items correctly identified out of 20 targets. On the computer mouse, Stroop, and Trail Making Tests, higher scores indicated poorer performance. On the Digit Span, Spatial Span and *n*-Back Tests, higher scores indicated better performance.

To allow for certain statistical analyses, the demographic variables were converted into dichotomous nominal data: Gender (male/female); order (Digit Span, Spatial Span, Trail Making, Stroop, *n*- Back/ *n*- Back, Stroop, Trail Making, Digit Span, Spatial Span); computer experience (experienced/not experienced); computer confidence (confident/not confident); reported anxiety (anxious/not anxious); home language (first language English/second language English); level of education (12-15 years/ 16-19 years); and English confidence (very confident/moderately confident). Some qualitative information may have been lost in the dichotomising of these variables. For instance, subjects had to respond whether they were very confident, moderately confident or not very confident speaking, reading and writing in English. However, none of the subjects reported *not* being confident speaking, reading and writing in English. The dichotomy consequently represented a division between subjects who reported being *very* confident and those who reported being *moderately* confident speaking, reading and writing in English. Similarly, second language English speakers could be further divided into first language Afrikaans, Zulu, Tswana, Xhosa, Sotho, Shona, German, and Siswati speakers. Further investigation of different languages and test performance is necessary for a more comprehensive understanding of this variable.

The results of the study are displayed in graphs and tables. Inferential and descriptive statistical techniques were used to summarise and analyse the data. Statistical analyses were conducted using SAS version 3.0. Outliers were removed and the data was subsequently presented in a final set of descriptive statistics. The suitability of the data for parametric analysis was explored. Correlational analysis of the effect of computer mouse performance on the tests was performed. Analyses of the demographic variables were performed to see whether these variables had an effect on performance on any of the tests. A factor analysis was then conducted in order to examine the construct validity of the *n*- Back Test, and to identify the nature of the relationship between the processes and components of working memory in the study.

3.2 Reliability

Issues of reliability are an important part of the development and utilisation of psychometric tests. The reliability of a test is defined as, "…the estimate of what proportion of variance in performance can be attributed to true differences in behaviour" (Gibertini & Retzlaff, 1994, p.187). Each measure of reliability looks at different parts of error. The term 'reliability' subsequently represents a number of concepts. These concepts include internal consistency, stability and inter-judge concordance. Internal consistency refers to the uniformity of the test and its individual items, and it examines the individual error that occurs in sampling behaviour. Stability refers to the degree to which a test measures a construct accurately across time. Inter-judge concordance refers, in the clinical context, to the degree to which two judges reach the same conclusion on a test that demands subjective scoring.

The reliabilities of the cognitive measures used in the present study have been well documented and established on the Stroop, Trail Making and Digit Span Tests (Lezak, 1995, 2004). Digit Span test-retest reliability coefficients vary between .66 and .89 (Lezak, 1995, 2004), reliability coefficients for the Trail Making Test vary between .60 and .90 (Lezak, 1995, 2004), and reliabilities on the interference score of the Stroop Test have been reported at .70 (Groth-Marnat, 2000). Due to the nature of the tests being used, often the only form of reliability that can be assessed is test-retest reliability. Test-retest reliability may be used to assess the stability of a neuropsychological test across time. This type of reliability could not be assessed on any of the tests within the time- frame of the study due to time constraints, resource constraints, and due to the fact that significant practice effects have been demonstrated in prior research on many of these tests. For example, practice effects on the Digit Span Test have been found to be moderately but statistically significant (Lezak, 1995). Practice effects have been found on the Trail Making Test on retesting, particularly on part A of the test (Lezak, 1995). Practice effects on the Stroop Test have also been demonstrated repeatedly (Groth-Marnat, 2000). It was subsequently not feasible to assess test-retest reliability for the present study. In future research, it would be both feasible and necessary to assess the test-retest reliability of the computer-based Stroop, Trail Making, Digit Span, Spatial Span, and *n*-Back Tests. Adequate time between administrations of the tests would be needed to counteract practice effects

The individual items on the tests used in this study had no intrinsic meaning, as the cognitive construct being assessed had more to do with the cognitive processes elicited by the test, than the test items themselves. This meant that there was no inter-item correlation on these tests. Cronbach's alpha measures inter-item correlation, and so it could not be used to measure reliability on the tests in this study. Additionally, Cronbach's alpha also does not adequately measure internal consistency in timed tests and so examination of the internal consistency of the computer mouse, Stroop and Trail Making Tests was not feasible (Gibertini & Retzlaff, 1994).

Split-half techniques could not be used to assess reliability in the computer mouse test, Stroop Test and Trail Making Test, as these tests were timed and did not have a cut-off point. Practice effects also occurred on these tests so that the test-taker speeded up from the first to the second half of each phase of the test. The tests that were not timed, including the Digit Span Test, the Spatial Span Test and the *n*- Back Test became more demanding towards the latter part of the test so that performance declined. Therefore, a comparison of performance on each half of the test was not appropriate. However, with the implementation of a cut-off point on the timed tasks, it may be both feasible and necessary to assess the split-half reliability of the speeded tests in future research. Alternate forms are another means by which reliability may be assessed. However, there were no accessible alternate versions of any of the tests used in the study. The creation of alternate forms for the tests may be an effective means for assessing the reliability of the tests in future research, particularly as the nature of the tests makes them difficult to assess in terms of other forms of reliability.

3.3 Removing outliers

Visual inspection of the data revealed that there were outliers in the sample. A conventional method of eliminating deviant responses further away than two standard deviations below and above the mean was first utilized. However, this method was found to eliminate 3.7 percent of the total data. In other words, 55 raw scores of a total of 1484 scores would be removed. The removal of this amount of data would detract from one of the purposes of the analysis, which was to explore individual variation in performance on the computer-based tests administered in the study. Furthermore, it was deemed unnecessary to remove responses that reflected above average performance. Subsequently, only responses that reflected significantly below average performance on each test were removed.

Removing outliers that were 2 or more standard deviations or more away from the mean only in the direction that indicated poor performance still eliminated 45 raw scores out of 1484, which was 3.0 percent of the total data. The fact remained that outliers that were abnormally far from the mean were problematic. Outliers that were extremely far from the mean could potentially confer a significant result to a statistical test when the rest of the scores would not (Aron & Aron, 1999). Consequently, responses that were three or more standard deviations away from the mean in the direction that indicated poor performance on each test were removed. This accounted for 12 raw scores, or 0.08 percent of the raw data, a much more acceptable figure. Only one raw score was removed from the computer mouse test. One raw score was removed from each of the word-reading and colour-naming phases of the Stroop Test. Four scores were removed from the 0-Back condition of the *n*- Back Test, 2 raw scores were removed from the 1-Back condition, and one raw score was removed from the 2-Back condition of the *n*- Back Test. One raw score was removed from each phase of the Trail Making Test. The basic statistics for each subtest after removal of outliers is displayed in Table 2.

Table 2.

Summary of performance of subjects on all the tests administered in the study

Variable	N	Mean	Median	Std. Dev	Variance	Minimum – maximum
Stroop1	105	87.50	86.28	10.12	102.61	65.86 - 116.92
Stroop2	104	80.14	79.49	9.03	81.58	60.8 – 103.19
Stroop3	104	91.26	88.74121.2320	13.11	171.99	68 – 128.47
Stroop4	105	123.02	18	18.33	336.07	82 - 168.27
0Back	105	20	15	-	-	-
1Back	105	18.01	0	1.82	3.33	12 - 20
2Back	105	14.37	7	3.09	9.55	5 - 20
3Back	105	6.35	6	6.82	46.55	0 - 18
Digits F	105	6.77	6	1.13	1.29	4 - 9
Digits B	105	5.61	5	1.22	1.50	2 - 9
Spatial F	105	5.84	31.92	1.28	1.64	4 - 9
Spatial B	105	5.28	51.04	1.17	1.36	3 - 9
Trails A	104	33.11	35.04	8.19	67.09	17.81 – 60.85
Trails B	104	51.32		11.36	129.10	25.68 – 72.92
Mouse Test	104	36.65		7.42	55.13	24.09 – 61.46

From Table 2, it can be seen that the mean time taken to complete the Stroop Test was 87.50 seconds on the word-reading phase (Stroop1), 80.14 seconds on the colour-naming phase (Stroop2), 91.26 seconds on the interference phase (Stroop3), and 123.02 seconds on the mixed interference phase of the test (Stroop4). The mean time decreased by 7.36 seconds from the word- reading (Stroop1) to the colour-naming (Stroop2) phase. It increased by 11.12 seconds from the Stroop2 to the interference phase (Stroop 3). It then increased further by 31.76 seconds from Stroop 3 to the mixed interference phase (Stroop4). The variance of the test decreased from Stroop 1 to Stroop 2, and increased considerably from Stroop 2 to Stroop 3 and from Stroop 3 to Stroop 4.

The mean number of items correctly identified out of 20 was 20 on the 0-Back condition of the *n*- Back Test, 18.01 on the 1-Back condition, 14.37 on the 2-Back condition, and 6.35 on the 3-Back condition of the *n*- Back Test. There was a mean drop of 1.99 points out of 20 from the 0-Back to the 1-Back condition, a mean drop of 3.64 points out of 20 from the 1-Back to the 2-Back condition, and a mean drop of 8.02 points out of 20 from the 2-Back to the 3-Back condition of the *n*- Back Test. The variance increased substantially from the 1- to the 3-Back conditions of the test.

The number of digits repeated in the correct sequence on the forward phase of the Digit Span Test was 6.77. This decreased by 1.16 digits in the backward (Digits B) condition. The variance increased slightly from the forward to the backward condition of the test. The mean number of items entered in the correct spatial sequence on the forward phase of the Spatial Span Test was 5.84, but this decreased by 0.56 items in the backward (Spatial B) condition. The variance also decreased slightly from the forward to the backward condition. The mean time taken to complete part A of the Trail Making Test was 33.11 seconds, but this increased by 18.21 seconds on part B (TrailsB) to 51.32 seconds. The variance also increased substantially from part A to part B of the Test.

3.4 Normality

Suitability of the data for parametric analysis was determined using a number of methods. Three important assumptions for parametric statistical analysis included random independent sampling, interval measurement scale, and normal distribution of test scores. First, although the sample was a haphazard convenience sample of volunteers, the order of the tests was randomly allocated to participants. The tests all utilised interval scales of measurement. The sample size was large enough (n=105) to assume normal distribution of test scores. Histograms for each subtest revealed a fairly normal distribution of raw scores on each test. The only subtests that were significantly skewed to the left indicated ceiling effects that reflected very good performance in the sample, as anticipated. The measures of central tendency indicated acceptable normality of data for the rest of the tests, particularly with regards to the distance between the median and the mean, demonstrated in Table 2 (Aron & Aron, 1999). The tests for a normal distribution are displayed in Appendix D.

Appendix D revealed that only the Stroop Test (Stroop1 and 2), and Trail Making Test (Trails A) were normally distributed with regards to the Kolmogorov-Smirnov statistic at $\alpha = .01$, a particularly sensitive measure of normality. The other tests did not demonstrate statistical significance on this measure. As the latter statistic was a very sensitive measure, normal distribution was also determined using measures of skewness and kurtosis (Aron & Aron, 1999). Skewness examines the symmetry in a distribution, with a normal distribution having a value close to 0. Kurtosis examines the peak in a distribution, or shape of the probability distribution, with a normal distribution having a value close to 0.

Examination of these statistics suggested that all the test data were relatively normally distributed, with the exception of the 0-Back and 3-Back conditions of the *n*- Back Test. However, floor effects on the 3-Back had been found in the literature and were therefore expected on this measure. The 0-Back condition did not have a distribution as, once the outliers had been removed, there was no variance on this test. The 0-Back condition of the *n*-Back Test was subsequently removed from further analysis. It was decided that, overall, the distributions of the remaining tests were appropriate to allow for parametric analysis.

3.5 Simple Correlations: Computer performance

The computer mouse test was correlated with each of the subtests in order to test the hypothesis that computer mouse performance would have a significant relationship with performance on speeded computerised tests. The study related computer mouse performance to performance on each of the tests. Correlational analysis was conducted using a two-tailed test at significance level $\alpha = .015$. A summary of the results is given in Table 3.

Table 3.

Correlations: Computer mouse test

	Pearson Correlation Coefficients		
Subtest	Mouse Test (*r*)	p- value	Degrees of freedom (*N*–2)
Stroop1	.60	p < .0001	102
Stroop2	.57	p < .0001	101
Stroop3	.49	p < .0001	101
Stroop4	.42	p < .0001	102
1Back	.02	p = .80	100
2Back	-.16	p = .10	101
3Back	-.15	p = .10	102
Digits Forward	-.15	p = .11	102
Digits Backward	-.16	p = .08	102
Spatial span	-.04	p = .64	102
Trails A	.44	p < .0001	101
Trails B	.28	p = .003	101

Table 3 revealed a large, significant, positive correlation between time taken to complete the computer mouse test and all the phases of the Stroop Test as well as part A of the Trail Making Test. A small but significant positive correlation was found between time taken to complete part B of the Trail Making Test and time taken to complete the computer mouse test. (*r* (101) = .28, p = .003). There were no significant correlation coefficients between computer mouse performance and the other tests in the study.

Cohen's convention of *r* = .10 for small, *r* = .30 for medium and *r* = .50 for large effect sizes was used (Aron & Aron, 1999). The study demonstrated medium effect sizes for the correlations between computer mouse performance and performance on Stroop1, Stroop2, Stroop3, Stroop4, Trails A at *r* = .60, *r* = .57, *r* = .49, *r* = .42, and *r* = .44, respectively. The power of the study to find a medium effect size was between .86 and 1.00, with *N* = 100 at the .05 significance level. However, the large variance on the Stroop and Trail Making Tests suggested that the power of the study was closer to .86 for a medium effect size. Part B of the Trail Making Test, with an effect size $r = .28$, and with large variance, revealed that the power of the study was closer to .17 to find a small effect size.

3.6 *t*-Tests

Independent sample t-tests (two-tailed) were conducted to test the hypotheses that there would be significant differences in performance between subjects based on computer experience, confidence using a computer, and test anxiety, and based on demographic variables such as gender, home language, level of education, and confidence speaking, reading and writing in English on the tests. There was evidence of equality of variance between most of the groups, so independent *t* tests for equal (or pooled) variances were conducted. However, there was no equality of variance between groups in which tests were administered in different orders on the computer mouse test. There was also no equality of variance between groups that reported being confident or not being confident using a computer on the computer mouse test. For these analyses, Satterthwaite's adjustment was used. To control for increases in type I error rate that occur when multiple *t*-tests are conducted, significance testing was conducted at $\alpha = .015$ (Harris, 1986). The results for each of the *t*-tests are summarised in Table 4.

Table 4.

Two-Sample Independent t-test Statistics

Subtest	Order	Gender	Home Language	Education	Reported Anxiety	Computer Confidence	Computer Experience	English Confidence
Stroop1	0.58	-2.83*	-2.70*	2.38	1.09	-1.95	-4.13**	-0.40
Stroop2	0.32	-3.04*	-2.69*	2.47	1.25	-1.69	-3.69**	-0.75
Stroop3	0.59	-1.62	-2.08	2.24	0.97	-2.73*	-3.16*	-0.92
Stroop4	0.12	-2.41	-2.43	1.37	0.76	-1.51	-2.56*	-2.17
1Back	-2.21	-0.56	-0.16	2.78 *	-0.38	-0.63	-1.24	0.99
2Back	-1.39	-0.49	1.88	2.24	-0.66	-0.09	-0.03	1.16
3Back	0.10	2.15	-1.02	0.18	1.08	0.05	-0.72	-1.10
Digits F	0.18	1.59	-0.69	0.57	-0.94	1.12	1.83	-0.96
Digits B	-0.36	1.09	2.63*	-0.91	-0.45	0.99	1.20	-0.22
Spatial F	1.19	1.23	0.51	2.34	-1.86	-0.61	0.64	-1.70
Trails A	0.15	-1.06	-3.27*	2.13	1.07	-1.80	-2.93*	-1.28
Trails B	1.51	-0.27	-2.96*	2.24	0.92	-0.82	-2.19	-1.13
Mouse Test	-0.30	-2.93 *	-2.46	0.53	-0.39	*-2.34*	-4.09**	0.34

Note: * = p<.01; ** = p<.001
Note: Satterthwaites' adjustment used in cases in which there was no equality of variance. Equality of variance was established at p > 0.01

There was a significant difference between males and females on time taken to complete the word-reading subtest of the Stroop Test (*t* = -2.83, df= 103, p <.015), the colour-naming subtest of the Stroop Test (*t* = -3.04, df= 102, p <.015), and the computer mouse test (*t* = -2.93, df= 102, p <.015). Males performed faster than females on these tests (Appendix E). A significant difference was found between subjects with 12-15 years of education and subjects with 16-19 years of education on 1-Back condition of the *n*-Back Test (*t* = 2.78, df= 101, p <.015). Surprisingly, subjects with 12 – 15 years of education correctly identified more items on the 1-Back condition of the *n*- Back Test than subjects with 16-19 years of education.

Significant differences were found between first and second language English speakers on the word-reading subtest of the Stroop Test (*t* = -2.70, df = 103, p < .015), on the colour-naming subtest of the Stroop Test (*t* = -2.69, df = 102, p < .015), and on part A (*t* = -3.27, df = 103, p < .015) and part B (*t* = -2.96, df = 103, p < .015) of the Trail Making Test. First language English speakers performed faster than second language English speakers on these subtests (Appendix E). There was a significant difference between first and second language English speakers on the backwards condition of the Digit Span Test (*t* = 2.63, df = 103, p <.015). First language English speakers had higher scores than second language English speakers on this test.

Significant differences were found between subjects who reported having had a lot of experience and those who reported not having had a lot of experience using a computer on the word- reading subtest (*t* = -4.13, df = 103, p < .001), on the colour-naming subtest (*t* = -3.69, df = 103, p < .001), on the interference subtest (*t* = -3.16, df = 103, p < .015), and on the mixed interference subtests of the Stroop Test (*t* = -2.73, df = 103, p < .015), as well as on part A of the Trail Making Test (*t* = -2.93, df = 103, p <.015), and on the computer mouse test (*t* = -4.09, df = 102, p <.001). Subjects who reported being experienced using a computer performed faster on all these tests than subjects who were not experienced using a computer (Appendix E). There was a significant difference between subjects who reported being confident using a computer and between subjects who reported not being confident using a computer on the interference subtest of the Stroop Test (*t* = -2.73, df = 102, p < .015). Subjects who reported being confident using a computer were faster than subjects who reported not being confident using a computer on this subtest (Appendix E).

There were no significant differences in performance on any of the tests according to the order in which the tests were administered, according to self-reported level of anxiety or according to self-reported confidence reading, writing and speaking in English. There were no significant differences in performance on the mixed interference phase of the Stroop Test, on the 2 - and 3 –Back conditions of the *n*- Back Test, on the forwards condition of the Digit Span Test or on the forwards condition of the Spatial Span Test according to gender, level of education, home language, computer experience, or computer confidence.

Effect Size

The effect size for significant differences between the groups was calculated using Cohen's delta (*d*) (Aron & Aron, 1999). Power was calculated at $\alpha = .05$ (Aron & Aron, 1999). The power of the study to find a small effect size was only .29. However, the power of the study to find a medium effect size was .94 and for a large effect size was close to 1. Medium effect sizes were found for the difference between males and females on the word-reading subtest ($d = -0.61$), and the colour-naming subtest of the Stroop Test ($d = - 0.67$), and on the computer mouse test ($d = - 0.68$). The means of the male group were at the 73rd, 75th, and 75th percentile of the female group, respectively.

Medium effect sizes were found for the difference between first and second language English speakers on the word-reading phase subtest ($d = - 0.56$), and the colour-naming subtest of the Stroop Test (.$d = - 0.58$), on the backward condition of the Digit Span Test ($d = 0.54$), and on part A ($d = - 0.65$) and part B ($d = - 0.64$) of the Trail Making Tests. The means of the first language English group were at the 71st, 72nd, 70th, and 75th percentile of the second language English group, respectively.

Large effect sizes were found for the difference between subjects who reported being experienced and subjects who reported not being experienced using a computer on the word-reading subtest ($d = - 0.90$), and the colour-naming subtest of the Stroop Test ($d = - 0.88$), and on the computer mouse test ($d = - 0.91$). The means of the experienced group were at the 82nd, 81st, and 82nd percentile of the non-experienced group, respectively. Medium effect sizes were found on the interference subtest of the Stroop Test ($d = - 0.69$), and on part A of the Trail Making Test ($d = 0.63$). The means of the experienced group were at the 76th and 74th percentile of the non-experienced group, respectively.

There was a medium effect size for the difference between subjects who reported being confident and subjects who reported not being confident on the interference phase of the Stroop Test ($d = 0.45$), and the mean of the confident group was at approximately the 67th percentile of the non-confident group. A small effect size was found for the difference between subjects with 12-15 years of education and subjects with 16-19 years of education on the 1- Back condition of the *n*-Back Test ($d = 0.34$). The mean of the group with 12-15 years of education was at the 63rd percentile of the group with 16-19 years of education.

3.7 Analysis of Covariance

The results of the *t*- tests suggested that there could be a confounding variable in the analysis that may have affected performance particularly on the speeded computer-based tests. It was proposed that performance on the computer mouse test may have acted as the confounding variable in the analysis. To test the assumption that computer mouse performance affected performance on the speeded tests, and could possibly account for the variance based on gender, education, home language and computer experience, a series of two-way Analyses of Covariance (ANCOVA) for unrelated measures were conducted with computer experience, home language, and level of education as the unrelated measures independent variables and with computer mouse scores as the covariate. The analysis was conducted to see whether the differences between subjects' scores based on computer experience, home language, and level of education on the Stroop and Trail Making Tests remained once the covariate had been removed. Significance testing was conducted at $\alpha = .05$. The results of the ANCOVA are presented in Table 5.

Table 5.

Analysis of Covariance

Subtest (Dependent Variable)	Degrees of Freedom	Computer Experience	Education	Home Language	Computer mouse Test
Stroop1	5, 99	F = 2.02	F = 4.05*	F = 0.22	F = 18.36***
Stroop2	4, 98	F = 1.48	F = 4.66*	F = 0.59	F = 16.02***
Stroop 3	4, 98	F = 0.68	F = 3.68	F = 0.02	F = 10.10***
Trails A	4, 98	F = 0.62	F = 2.01	F = 2.17	F = 8.50***
Trails B	4, 98	F = 0.22	F = 1.85	F = 3.05	F = 4.45**

Note: * p < .05, ** p < .01** *p < .001

The analysis showed no significant main effects of computer experience or home language on any of the tests. The analysis showed a significant effect of level of education on the word-reading phase of the Stroop Test, (*F* (5, 99) = 4.05, p = .04). Subjects with 12-15 years of Education were faster than subjects with 16-19 years of Education on the test (adjusted *M* = 89.43; adjusted *M* = 85.45, respectively) The covariate was significantly related to the dependent variable (*F* (5, 99) = 18.36, p < .0001), indicating that both computer mouse ability and level of education were related to performance on the word-reading phase of the Stroop Test.

The analysis showed a significant effect of level of education on the colour-naming phase of the Stroop Test, (*F* (4, 98) = 4.66, p = .03). Subjects with 12-15 years of Education were faster than subjects with 16-19 years of Education on the test (adjusted *M* = 82.00; adjusted *M* = 78.10, respectively). The covariate was significantly related to the dependent variable, (*F* (4, 98) = 16.02, p < .0001), indicating that both computer mouse ability and level of education were related to performance on the colour-naming phase of the Stroop Test.

For the rest of the analysis, there were no significant effects of computer experience, home language or level of education on any of the tests once computer mouse performance was partialed from the analysis. Only the covariate was significantly related to the interference phase of the Stroop Test (*F* (4, 98) = 10.10, p < .0001), to part A of the Trail Making Test (*F* (4, 98) = 8.50, p < .0001), and to part B of the Trail Making Test (*F* (4, 98) = 4.45, p = .002).

3.8 Correlation Matrix

The correlation matrix was inspected to investigate the relationships between the some of the variables of the study. The correlation matrix is reported in Appendix F. The matrix revealed strong significant correlations between performance on the computer mouse test and performance on the speeded Stroop and Trail Making Tests, which suggested that computer mouse performance affected performance on some of these tests. The computer mouse test scores were subsequently partialed from the correlation matrix. Inspection of an initial matrix revealed that the backward phase of the Spatial Span Test did not correlate significantly with any other variable in the matrix. The backward phase of the Spatial Span Test was subsequently removed from the analysis. Every remaining variable was significantly correlated with at least one other variable in the matrix. The correlation matrix partialing computer mouse performance is reported in Table 6. Significant values for the partial correlation matrix were determined at $\alpha = .015$ (Kanji, 1993).

Table 6

Partial Correlation Matrix

Subtest	Stroop 1	Stroop 2	Stroop 3	Stroop 4	1 Back	2 Back	3 Back	Digits F	Digits B	Spatial F	Trails A	Trails B
Stroop1	1.00	.773**	.728**	.587**	.037	-.070	.016	-.094	-.146	-.019	.431**	.330*
Stroop2	.773**	1.00	.791**	.639**	-.029	-.118	-.054	-.043	-.052	-.134	.366*	.314*
Stroop3	.728**	.791**	1.00	.711**	-.103	-.106	.054	-.011	-.086	-.068	.336*	.275*
Stroop4	.587**	.639**	.711**	1.00	-.054	-.242	-.031	-.087	-.183	-.114	.329*	.367*
1 Back	.037	-.029	-.103	-.054	1.00	.552**	.235	-.024	.127	.125	.006	-.118
2 Back	-.070	-.118	-.106	-.242	.552**	1.00	.279*	.150	.315*	.248*	-.200	-.306*
3 Back	.016	-.054	.054	-.031	.235	.279*	1.00	-.049	.043	.165	-.052	-.124
DigitsF	-.094	-.043	-.011	-.087	-.024	.150	-.049	1.00	.425**	.215	-.185	-.110
DigitsB	-.146	-.052	-.086	-.183	.127	.315*	.043	.425**	1.00	.117	-.196	-.088
SpatialF	-.019	-.134	-.068	-.114	.125	.248*	.165	.215	.117	1.00	-.084	-.130
TrailsA	.431**	.366*	.336*	.329*	.006	-.200	-.052	-.185	-.196	-.084	1.00	.565**
TrailsB	.330*	.314*	.275*	.367*	-.118	-.306*	-.124	-.110	-.088	-.130	.565**	1.00

Note: * $p < .015$; ** $p < .0001$

The four phases of the Stroop Test were highly correlated with one another. The subtests of the Stroop Test correlated significantly with parts A and B of the Trail Making Test: the word-reading sub-test correlated significantly with part A of the Trail Making Test (r (97) = .43, $p < .0001$) and with part B (r (97) = .33, $p = .001$). The colour-naming subtest correlated significantly with part A of the Trail Making Test (r (97) = .36, $p = .0002$) and with part B (r (97) = .31, $p = .001$). The interference subtest correlated significantly with part A (*r* (98) = .33, p = .0008) and weakly but significantly with part B (*r* (98) = .27, p = .006). The mixed interference subtest correlated moderately and significantly with part A (*r* (99) = .32, p = .0008) and with part B (*r* (99) = .36, p = .0002).

The 2- Back subtest of the *n*- Back Test correlated significantly with a number of variables. It correlated significantly with the 1-Back subtest (*r* (99) = .55, p< .0001), and it correlated weakly but significantly with the 3-Back subtest (*r* (101) = .27, p = .003). The 2-Back subtest correlated significantly with Part B of the Trail Making Test (r (101) = -.30, p = .002), and with the backwards condition of the Digit Span Test (*r* (101) = .31, p = .001). The 1- and 3-Back subtests of the *n* – Back Test only correlated with the 2- Back condition of the Test, but did not correlate with any other variable, or with one other (*r* (99) = .23, n.s). Contrary to expectation, the 2-Back condition correlated weakly but significantly with the forward condition of the Spatial Span Test (*r* (101) = .24, p = .012), and did not correlate with the mixed interference phase of the Stroop Test (*r* (99) = -.24, n.s.).

Part A and B of the Trail Making Test were highly correlated with one another (*r* (102) = .56, p < .0001). The forward and backward conditions of the Digit Span Test were also significantly correlated (*r* (102) = .42, p < .0001). The remainder of the partial correlation matrix revealed non-significant correlations between the variables. The low partial correlation matrix indicated that there was collinearity amongst the variables. This indicated that there could be hidden or latent relationships between the variables of the study that could be examined using factor analysis. An exploratory factor analysis was conducted to investigate the latent relationships between the variables of the study.

3.9 Factor Analysis

This study conducted an Exploratory Factor Analysis (EFA) in order to determine the factor structure underlying the variables in the correlation matrix (Costello and Osbourne, 2005). The primary aim of EFA in this study was to investigate the construct validity of the N-back Test, by comparing performance on this test to other tests that have been theoretically linked to working memory (Kane et al., 2007). Factor analysis was used to test the hypothesis that the *n*-Back Test would correlate with the backward condition of the Digit Span Test and the backward condition of the Spatial Span Test as measures of working memory span. EFA is a statistical method of accounting for common variance in a set of variables, and is frequently used as a method of multivariate analysis in the social sciences (Costello and Osbourne, 2005). Factor analysis has been put forward as a good multivariate method for examining convergent and divergent construct validity (Pulos, 1997).

3.9.1 Removal of variables

There was no variance on the 0-Back condition of the *n*-Back Test. This variable was subsequently removed from statistical investigation. Kaiser's Measure of Sampling Adequacy (MSA) was examined to determine whether the correlation matrix was suitable for analysis. The correlation matrix had revealed that the backward phase of the Spatial Span Test did not correlate significantly with any other variable in the matrix. Removal of this variable from the correlation matrix raised the MSA from .74 to .75. Overall MSA for analysis was acceptable at .75. No further variables were removed. The method of extraction used for the factor analysis was the principal components method.

3.9.2 Number of factors retained for analysis

The number of factors to be retained for analysis was based on a number of criteria. Kaiser's criterion that eigenvalues greater than 1.0 indicated common factors and should therefore be retained was acknowledged. However, this method was recognised as a method that consistently over-estimated the number of factors to be included in the analysis (Costello & Osbourne, 2005; Kline, 1994). The eigenvalues of the correlation matrix are presented in Table 7.

Table 7.

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	3.77	1.80	0.31	0.31
2	1.97	0.60	0.16	0.47
3	1.37	0.31	0.11	0.59
4	1.05	0.11	0.08	0.68
5	0.93	0.16	0.07	0.75
6	0.77	0.24	0.06	0.82
7	0.52	0.04	0.04	0.86
8	0.48	0.11	0.04	0.90
9	0.37	0.03	0.03	0.93
10	0.34	0.13	0.02	0.96
11	0.20	0.03	0.01	0.98
12	0.16		0.01	1.00

Table 7 demonstrated that four eigen values exceeded 1.0. The first was substantial, at 3.77, and explained 31 percent of the common variance. The second, at 1.97, explained 16 percent of the common variance, and 47 percent of the cumulative variance. The third, at 1.37, explained 11 percent of the common variance, and 59 percent of the cumulative variance. The fourth, at 1.05, explained 8 percent of the common variance, and 68 percent of the cumulative variance. The results suggested that the selection of three or more factors for the analysis would explain a sufficient proportion of the common variance of the correlation matrix.

Cattell’s scree plot was utilised as another method for selecting the number of factors to retain for analysis. This method is based on the graphical representation of the eigenvalues, and is a fairly subjective method for determining the number of factors to be retained (Kline, 1994). The graph was analysed according to a criterion whereby the point at which the curve “flattens out” determined the cut-off point above which the number of eigenvalues were retained as factors and below which the remaining eigenvalues were rejected (Costello and Osbourne, 2005). Cattell’s scree plot of the eigenvalues is presented in Appendix G.

Cattell’s scree plot of the eigenvalues demonstrated that the line changed slope after three eigenvalues. Selection of the number of factors to be retained for the analysis was based on a combination of methods. The eigenvalues of the correlation matrix demonstrated that there were four eigenvalues that were greater than 1. However, the first three eigenvalues each explained a significant portion of the variance whilst the rest of the values explained similar smaller portions of variance. The cumulative percentage of variance explained by three factors was good at 59 percent. Finally, Cattell’s scree plot revealed that the line changed slope after the cut-off point of three eigenvalues. Three factors were subsequently retained for the analysis (Kline, 1994).

3.9.3 Unrotated Factor Analysis

The factor pattern for retention of three factors is presented in Table 8. Loadings of .30 have been put forward as a significant minimum loading for well-defined factors in the literature (Kline, 1994; Costello & Osbourne, 2005). Due to the fact that this study had a sample size of only $N = 105$, and that the factors were moderately well-defined, significant loadings were determined more conservatively at a cut-off item loading of .40.

Table 8

Factor pattern for the retention of three factors

Subtest	Factor1	Factor2	Factor3
Stroop1	.81	.32	-.01
Stroop2	.83	.29	.13
Stroop3	.82	.31	.14
Stroop4	.79	.13	.06
1Back	-.17	.61	-.46
2Back	-.37	.73	-.19
3Back	-.10	.46	-.41
Digits F	-.21	.33	.74
Digits B	-.29	.44	.54
Spatial F	-.22	.41	.08
Trails A	.62	-.06	-.21
Trails B	.58	-.18	.01

Table 8 revealed that simple structure was not obtained on an unrotated factor pattern. The word-reading (Stroop1), colour-naming (Stroop2), interference (Stroop3) and mixed interference (Stroop4) phases of the Stroop Test loaded significantly onto the first factor at .81, .83, .82 and .79, respectively. Both conditions of the Trail Making Test loaded significantly onto the first factor at .62 and .58, respectively. The 1-Back condition of the *n*-Back Test loaded significantly onto the second and third factor at .61 and .46, respectively. The 2-Back condition of the *n*-Back Test loaded significantly onto the second factor at .73, and the 3-Back condition of the *n*-Back Test loaded significantly onto the second and third factors at .46 and .41, respectively. The forward condition of the Digit Span Test (Digits F) loaded significantly onto the third factor at .74. The backward condition of the Digit Span Test (Digits B) loaded significantly onto both the second and third factors at .44 and .54, respectively. The forward condition of the Spatial Span Test (Spatial F) loaded significantly onto the second factor at .41.

3.9.4 Rotated factor analysis

Kline (1994) noted that it was often necessary in a factor analysis to rotate the factor in order to reveal a clearer or simpler factor structure. Rotation was done to examine the same, shared variance between the items from a different position and to examine whether the rotated factor pattern better explained the variance. Varimax orthogonal rotation was selected in order to examine the factors, under the assumption that the factors were uncorrelated with one another (Costello & Osbourne, 2005). The results of the factor pattern for an orthogonal rotation are presented in Table 9.

Table 9.

Orthogonal Varimax Rotation

Subtest	Factor1	Factor2	Factor3
Stroop1	.87	.09	-.06
Stroop2	.89	-.02	.04
Stroop3	.89	-.00	.06
Stroop4	.80	-.10	-.08
1Back	-.00	.79	-.02
2Back	-.13	.78	.30
3Back	.00	.62	-.08
Digits F	-.01	-.09	.83
Digits B	-.07	.12	.74
Spatial F	-.06	.33	.33
Trails A	.54	-.06	-.36
Trails B	.49	-.28	-.23

Table 9 revealed fairly simple structure on a rotated factor analysis. There were as many zero loadings, or close to zero loadings on each factor as there were factors in the factor loading matrix. Each variable also only loaded significantly onto one factor. The word-reading, colour-naming, interference and mixed interference phases of the Stroop Test loaded significantly onto the first factor at .87, .89, .89, and .79 respectively, and not onto any other factor. Both part A and part B of the Trail Making Test loaded significantly onto the first factor at .54 and .49, respectively. The 1-, 2-, and 3-Back conditions of the *n*-Back Test loaded significantly onto the second factor at .79, .78, and .62 respectively, and not onto any other factor. Both the forward and backward conditions of the Digit Span Test loaded significantly onto the third factor at 0.83 and 0.74 respectively, and not onto any other factor. The forward condition of the Spatial Span Test loaded equally, but not significantly onto the second and third factors.

The first factor was comprised of significant loadings on Stroop1, Stroop2, Stroop3, Stroop 4, and parts A and B of the Trail Making Test. The second factor was comprised of significant loadings on the 1- 2- and 3-Back conditions of the *n*-Back Test. The third factor had significant loadings on the forward and backward phases of the Digit Span Test. The residual matrix was examined to estimate the quality of the factor analysis (Appendix H). The items in the residual correlation matrix should be close to 0 in order to determine how precisely the correlations can be replicated from the factors (Kline, 1994). The residual matrix for the factor analysis demonstrated low residual correlations, with most items in the residual matrix close to 0.

The results demonstrated that computer mouse performance was significantly correlated with the Stroop Colour Word and Trail Making Tests, and that computer mouse ability was also related to prior experience using a computer, confidence using a computer, gender and home language. An Analysis of Covariance (ANCOVA) was subsequently performed with computer mouse ability partialed out of the analysis as a covariate. No significant main effects of computer experience, confidence, gender or home language were found when computer mouse ability was removed from the analysis. Once computer mouse ability had been partialed from the results, the study found that the 2-Back condition of the *n*-Back Test correlated significantly with the backward condition of the Digit Span Test, the forward condition of the Spatial Span Test, and part B of the Trail Making Test. However, the *n*-Back Test did not load onto the same factor as these tests. Factor analyses revealed that the Stroop and Trail Making Tests loaded onto one factor, the *n*-Back Test loaded onto a second factor, and the Digit Span Test loaded onto a third factor.

CHAPTER FOUR

Discussion

4.1 Introduction

Neuropsychological testing has become increasingly important in the South African context where the incidence of traumatic brain injury is very high, and where neurocognitive impairment has been found in a number of diseases and organic brain disorders (Foxcroft, 1997; Foxcroft & Roodt, 2005). The functional disability experienced by those with organic brain trauma and disease has often not been proportional to damage evidenced in neuroimaging, electrophysiological and other brain imaging techniques. Neuropsychological testing has become an important means by which impairment in cognitive-behavioural functioning may be detected. The relatively recent introduction of computer-based tests has produced a new set of challenges in the testing context, particularly in South Africa where computer literacy is uncommon in the general population (Davies et al., 2005).

Working memory has been implicated in the breakdown of cognitive-behavioural functioning in a number of organic and psychiatric brain disorders (McAllister, 2006a, 2006b). The incidence of working memory deficits in traumatic brain injury, Alzheimer’s disease, Parkinson’s disease, Schizophrenia, Multiple Sclerosis, HIV and other disorders has made it an important and relevant concept in international neuroimaging studies, in modern experimental psychology, and in the South African context. This study investigated South African adults’ performance on a selection of computer-based neuropsychological tests of working memory.

Three key research questions guided and directed the theoretical development of the study. The first question aimed to explore key issues around computer-based testing in the South African context. The rationale for this question was anchored in research that suggested that computer literacy, familiarity using a computer, and test anxiety could affect performance on computer-based tests. Some research had suggested that even the ability to operate a computer mouse could affect test performance (Davies et al., 2005; Fisher, 2006). The recent introduction of international and local guidelines for computer-based test development and use has established the need for exploring administrative and psychometric issues in computer-based testing. The current study aimed to outline some of these concerns to guide the further development of the tests used in the study, and to contribute to the literature on computer-based neuropsychological testing in South Africa.

The first research question also examined the relationships between demographic variables such as gender, age, level of education, language, and performance on the computer-based tests. This question aimed to highlight issues of bias and fairness in computer-based neuropsychological testing, to explore whether computer-based administration of standardised neuropsychological tests affected the validity of the tests, and to explore the generalisability of the findings of this study (Foxcroft, 1997). The current study acknowledged that further investigation in different cultural, linguistic and social groups was necessary for a holistic appreciation of issues of bias and fairness in South Africa (Davies et al., 2005).

The second research question enquired whether the tests utilised in the study would reveal domain-specific relationships within each of the verbal, visual and spatial subcomponents of working memory, respectively, or whether the study would reveal stronger relationships between the tests according to domain-free executive processes. This question stemmed from the fact that these tests had been used extensively to measure short-term and working memory, and executive attention. However, the relationships between these tests had not been investigated previously in the literature. Previous studies tended to examine either the storage or attentional facets of working memory tasks, but did not often attempt to look at both these aspects of working memory together (Carpenter et al., 2000; Conway et al., 2005; Wager & Smith, 2003). The current study aimed to consider how currently used tests of working memory could be broken into different levels of processing and how these levels operated together. It also aimed to consider whether working memory should be assessed using a number of tests, or whether only one test was necessary to measure working memory.

The third research question enquired how the *n*-Back Test related to other tests of executive attention, short-term and working memory capacity. The *n*- Back Test had been used extensively in international neuroimaging research to investigate the construct of working memory (Owen et al., 2005). A small body of research on the validity of the *n*- Back Test had produced contradictory findings. Some research suggested that the *n*- Back Test was not a valid measure of working memory (Kane et al., 2007). Other research suggested that the *n*- Back Test had good face and content validity as a measure of working memory (Owen et al., 2005). These contradictory findings highlight the need to further investigate this topic, so the third research question was posed to investigate the validity of the *n*- Back Test (Conway et al., 2005; Kane et al., 2007).

The three general questions of the study gave rise to a number of hypotheses that were investigated utilising a variety of statistical procedures. The statistical results are discussed and interpreted, in turn, in relation to each of the research questions and hypotheses of the study. The limitations of the study and the key implications of the findings are outlined, followed by a discussion of directions for future research.

4.2 Results and Interpretation

4.2.1 Computer-based testing in the South African context

Computer-based testing has been used to enhance the administration of cognitive and neuropsychological tests. There are many advantages to computer-based test administration including level of enjoyment, speed and flexibility of administration, standardisation, accuracy in timing, reduced data handling error, and increased security of test results (Davies et al., 2005; Gur et al., 2001). The recent introduction of computer-based testing to the South African testing community has raised a number of concerns, particularly with regard to lack of exposure to information technology and varying levels of computer literacy in the population (Davies et al., 2005). Questions have been raised around the relationships between computer-based test performance, the ability to operate a computer mouse, test anxiety, confidence or experience using a computer, and social, cultural and linguistic factors (Davies et al., 2005; Fisher, 2006).

To address the first research question of the study, and to examine some of these issues, a number of statistical methods were employed. First, the computer mouse test was correlated with each of the subtests in order to test the hypothesis that computer mouse performance would have a significant relationship with performance on speeded computerised tests. Second, a series of two sample independent *t*-tests were conducted to test the hypotheses that there would be significant differences in performance between subjects based on computer experience, confidence using a computer, and test anxiety, and based on demographic variables such as gender, home language, level of education, and confidence speaking, reading and writing in English.

Not surprisingly, the results revealed that computer mouse performance was significantly correlated with all the subtests of the Stroop Test and with parts A and B of the Trail Making Test, both speeded tests (Table 3). The results suggested that the speed with which one manipulates a computer mouse may affect performance on speeded tests, and that subjects who operated a computer mouse more efficiently may have had an advantage over subjects who were slower on the computer mouse. This finding was consistent with the hypothesis that computer mouse ability would affect performance selectively on speeded computer-based tests.

The t-tests revealed a significant difference between subjects who reported being confident and not being confident using a computer on the interference subtest of the Stroop Test (t = -2.73, df = 102, p < .015). Significant differences between males and females, and first and second language English speakers, on the word-reading and colour-naming subtests of the Stroop Test were found (Table 4). There was a significant difference between males and females on the computer mouse test (t= -2.93, df = 102, p < .015). Significant differences were found between subjects who reported having had experience and not having had experience using a computer on the word- reading, colour-naming, and interference subtests of the Stroop Test, on part A of the Trail Making Test, and on the computer mouse test (Table 4). Subjects who were confident or experienced using a computer were faster than subjects who were not confident or experienced using a computer, males were faster than females, and first language English speakers were faster than second language English speakers on these tests. These findings were consistent with the hypotheses that computer experience, and confidence using a computer would affect performance on computer-based tests and that there would consequently be some artificial differences in performance according to gender, home language, and other demographic variables on the tests.

Examination of the raw data supported the hypothesis that the differences between the groups based on home language, computer experience and gender may have been due to the effect of a third variable on performance, rather than reflect differences in cognitive ability between the groups. The raw data revealed that approximately 32% of the female sample reported being either inexperienced or unconfident using a computer while only 23% of the male sample reported being either inexperienced or unconfident. It also demonstrated that approximately 50% of second language English speakers reported being inexperienced or unconfident using a computer while only 24% of first language English speakers reported being inexperienced or unconfident using a computer, and that approximately 4% of subjects with 16-19 years of education reported being inexperienced or unconfident using a computer, while 37% of subjects with 12-15 years of education reported being inexperienced or unconfident using a computer.

The raw data suggested that computer experience may have confounded performance particularly on the speeded, computer-based tests. The fact that the differences between males and females, first and second language English speakers, and subjects with 12-15 or 16-19 years of education were only on the subtests that had a large speed component and a smaller cognitive component supports this view. The difference in performance between males and females on the first two phases of the Stroop Test could reflect a difference in terms of exposure to information technology at home and at school. It could also reflect a socio-cultural gender difference in relation to access to, interest in, and training in computer use. Due in part to historical disadvantages, the difference between first and second language English speakers could indicate that many second language English speakers may not have had adequate access to computers or to information technology at home and at school (Davies et al, 2005). Subjects with higher levels of education may have had more exposure to computers, particularly in tertiary institutions where computer use is necessary predominantly at a postgraduate level.

The t-tests revealed a significant difference between subjects who were experienced and subjects who were not experienced using a computer on the computer mouse test (t = -4.09, df = 102, p <.001). It was proposed that computer familiarity, measured by performance on the computer mouse test and informed by experience using a computer, may have confounded performance on some of the tests. Evidence for a confounding effect of computer familiarity on test performance came from a number of sources. First, the significant correlations between computer mouse performance and the Stroop and Trail Making Tests suggested that the ability to operate a computer mouse, and the speed with which one could do so, was strongly related to performance on the speeded tests. Previous qualitative research also suggested that there was a relationship between computer mouse ability and performance on a selection of speeded computerised cognitive-neuropsychological tests (Fisher, 2006). Some research suggested that the introduction of practice tasks, and familiarisation of the subject with the requirements of computer-based tests, could lessen the impact of computer literacy on test performance (Davies et al., 2005). However, the introduction of practice tasks has not always improved test performance. In this study, each of the subtests was only administered once a practice task had been administered and completed successfully. The results suggested that performance on the computer mouse test continued to affect performance on the speeded Stroop and Trail Making Tests, even after the subject had familiarised him/herself with the test requirements through practice.

To test the assumption that computer mouse performance selectively affected performance on the speeded tests, and could account for the variance caused by gender, education, home language and computer experience as a product of computer familiarity, series of two-way Analyses of Covariance (ANCOVA) were conducted. The analysis was performed to see whether the differences between subjects' scores based on computer experience, home language, and level of education on the speeded Stroop and Trail Making Tests remained once the covariate had been removed. The analysis revealed no significant main effects of computer experience, gender, or home language on any of the tests once the covariate had been removed. However, the covariate was significantly related to the word-reading, colour-naming, and interference subtests of the Stroop Test, and to parts A and B of the Trail Making Test (Table 5).

The results of this analysis suggested that computer mouse performance, or computer ability, acted as a confounding variable, and that the ability to operate a computer mouse was influenced by a combination of social, historical and culturally-influenced factors such as home language, level of education, and experience and confidence using a computer. The adjusted means revealed that there were no differences in test performance between first and second language English speakers, between subjects with 12 to 15, and 16 to 19 years of education, and between subjects who reported being experienced or not experienced using a computer, on any subtest once computer mouse performance was partialed from the analysis. These results indicate that computer-based assessment may be more valid if a baseline test of computer mouse ability is performed and taken into account when analysing test results, particularly on speeded computer-based tests.

Additional findings

The t-tests also revealed a significant difference between first and second language English speakers on the backward condition of the Digit Span Test ($t = 2.63$, $df = 103$, $p < .015$). First language English speakers recalled more digits in the correct order than second language English speakers on this test. A possible explanation for this difference was that second language English speakers may not have understood the requirements for this part of the test as clearly as first language English speakers. This interpretation may be supported by the fact that the backward phase of the Digit Span Test did not have a practise trial and only written instructions were given for this more complicated phase of the test. The instructions for the backward phase may not have been clear or detailed enough, particularly to second language English speakers. However, almost 91% of the sample reported being very confident speaking, reading and writing in English, only 9% of the sample reported being moderately comfortable, and no subjects reported not being comfortable communicating in English.

On the other hand, Grieve (2005) reported that language was the most significant factor that affected psychometric test results in the South African context, and that it took longer for a second- language English speaker to understand and handle information that was presented in English, no matter how confident a person felt about communicating in English. Grieve (2005) also maintained that subjects who were second language English speakers but had been educated in English and had studied at an English-medium tertiary institution could have a double disadvantage in that they did not usually have a comparable advantage to first language English speakers, and their ability to speak, read and write in their home language could also be affected.

An alternate explanation for the difference between first and second language English speakers on the backward condition of the Digit Span Test is that the type of response required by the subject on the Digit Span Test may have added unnecessary complexity to the task, particularly for subjects unfamiliar with information technology tools (Gur et al., 2001). The fact that almost 50% of second language English speakers in this study reported being inexperienced using a computer could support this view. However, it is likely that performance on the forward phase of the test would also have been affected if this were the case.

There was a significant difference between subjects with 12-15 years of education and subjects with 16-19 years of education on the 1-Back condition of the *n*-Back Test ($t = 2.78$, df= 101, $p < .015$). Surprisingly, subjects with 12-15 years of education identified more correct stimuli on this task than subjects with 16-19 years of education. The author could not find an appropriate answer for this finding other than to suggest that perhaps the undergraduate students were better motivated to perform on the test. Alternatively, this result could have been a chance result, as reflected in the power of the study at approximately 29% for a small effect size. The power of this finding was not high enough to rule out the possibility that this result occurred by chance.

There were no significant differences in performance on any of the tests according to reported anxiety, or to the order in which the tests were administered (Table 4). These results suggest that there were no order effects, and that fatigue was not a factor in performance on these tests. The fact that there were no significant differences between subjects who reported being anxious before testing and those who reported not being anxious may be an indication that the practise trials on the tests allowed anxious subjects to familiarise themselves with the tests and may subsequently have lowered their level of anxiety. Davies et al., (2005) noted that assessment anxiety could be reduced when the element of surprise was removed from the testing context and in the case that the test administrator encouraged and supported the test-taker.

The finding that computer mouse performance acted as a covariate in the study on the speeded computer-based tests, and that it accounted for the variance in performance due to education, home language and computer experience on these tests suggests that partialing computer mouse performance from performance on the speeded Stroop and Trail Making Tests in future research may account for the influence of socio-culturally and historically influenced computer familiarity on computer-based test performance. Computer mouse performance was partialed from further analyses, in order to remove this effect from performance on the tests in this study.

4.2.2 Components and processes of working memory

To investigate the relationships between tests that tapped into the storage capacities and processing abilities of working memory, analysis of the correlational relationships between the tests was performed. Each of the memory tests measured the ability to store and recall verbal, spatial or visual information. Tests that theoretically tapped into working memory included an additional requirement in that the subject retained information short-term whilst simultaneously switching attention to another task (Conway et al., 2005). Tests that did not include a memory component but tapped into the ability to utilise complex executive attention were correlated with tests of short-term and working memory. It was proposed that the relationships between the tests would reveal whether there were domain-specific relationships within the tests, according to verbal, visual and spatial domains, or domain-free relationships between the tests, according to the use of complex executive attention.

Neuroimaging evidence had supported a distinction between verbal, visual and spatial subsystems in the brain. Neuroimaging studies found that verbal working memory tasks utilised a selection of areas predominantly in the left hemisphere of the brain, whereas visuospatial working memory tasks generally utilised areas predominantly in the right hemisphere (Baddeley, 1996c; Gazzaniga et al., 2002). Visual and spatial components were also localised to different areas of the brain (Nyberg & Cabeza, 2000). On the other hand, neuroimaging research also found evidence for integration of areas of the brain according to the use of domain-free complex executive attention. The Dorsolateral Prefrontal Cortex (DLPFC) was consistently activated in tasks in which an item was held over a delay and whilst the item was manipulated, across a number of domains (Callicott et al., 1999). Neuroimaging research also found that the same prefrontal areas of the brain were activated when tasks were *combined* whilst individual completion of the tasks activated domain-specific areas of the brain (Baddeley, 2000a).

It was proposed that the correlational analysis would reveal which tests or subtests tapped into domain-specific components of working memory, and which subtests tapped into domain-free executive attention. Having carefully reviewed the literature, it was proposed that tests of executive attention that did not include a memory component would have a significant relationship with working memory tests that incorporated central executive processes. The study hypothesised that the *n*- Back Test and the backward conditions of the Digit Span and Spatial Span Tests would correlate with part B of the Trail Making Test, and with the interference and mixed interference phases of the Stroop Test according to the use of domain-free complex executive attention.

An initial examination of the correlation matrix revealed that the backwards phase of the Spatial Span Test did not correlate significantly with any other variable in the matrix. This finding was unexpected, as the backward condition of the Spatial Span Test had been put forward as a measure of working memory capacity in the Wechsler Adult Intelligence Scale (WAIS –III) and Working Memory Scales (WMS – III). Additionally, the Spatial Span Test had loaded significantly onto the working memory indices of the WMS –III (Ponsford, 2000). On the other hand, a small body of research suggested that the backward phase of the Spatial Span Test may not be a valid measure of working memory span. Wilde, Strauss and Tulsky (2004) re-examined the WAIS-III and WMS-III standardisation and clinical group data and found that performance on the backward phase of the Spatial Span Test was at least as good as, or better than, performance on the forward condition of the test in 34.5% of the standardisation sample.

Wilde et al. (2004) proposed a number of reasons for this finding. First, they contended that as the same sequences were given on both the forward and backward conditions of the WAIS-III Spatial Span Test, subjects saw each sequence for a second time and so a practice effect may have occurred from the forward to the backward phase of the test. An alternate explanation was that the Spatial Span Test may not be an adequate test of working memory span, as it required subjects to recall the order or sequence of presentation of the items but not to remember the items themselves. Wilde et al. (2004) cited research that found that recall of only the order of items resulted in similar or equal forward and backward spans. They hypothesised that the backward phase of the Spatial Span Test did not involve further executive resources for active manipulation and reversal of the spatial items in working memory.

The results of this study almost replicated those of Wilde et al. (2004). This study revealed that performance on the backward phase of the Spatial Span Test was at least as good as performance on the forward phase of the test in approximately 47% of the sample. The raw data revealed that 28 subjects (26.6% of the sample) performed better on the backward than the forward phase of the test. There was no difference in performance on the forward and backward conditions of the Spatial Span Test in 22 subjects (20.9% of the sample). Additionally, the mean difference between the forward condition (*M* = 5.84) and the backward condition (*M* = 5.28) was negligible.

There was also a very low, non-significant correlation (r = .05, n.s.) between the forward and backward conditions of the Spatial Span Test. The low correlation suggested that these subtests did not tap into the same cognitive construct. One explanation is that there may have been a practice effect from the forward to the backward condition of the test. Support for this explanation comes from the fact that the variance decreased slightly from the forward to the backward condition of the Spatial Span Test (Table 2). This suggests that the backward phase may have been less demanding than the forward phase of the test. Different sequences were given on the forward and backward conditions of the Spatial Span Test in this study and so a practice effect would not have occurred for this particular reason. However, a practice effect may have resulted from better understanding of, and acquaintance with, the demands of the test on the backward condition once the subject completed the forward condition of the test. Better control of the computer mouse on the backward condition of the test may also have produced a practice effect. Had there not been a practice effect, the forward and backward conditions would hypothetically have measured the same cognitive construct, and there should have been a large, significant correlation between the subtests.

A recent neuroimaging study suggested a bias in the amount of central executive resources demanded by the visuospatial sketchpad (Rudkin et al., 2007). The study found that serial sequential processing of spatial information placed greater demand on the areas of the brain responsible for central executive functioning than simultaneous visuospatial processing, or than serial recall of verbal information. Rudkin et al. (2007) also reported neuroimaging evidence that supplemented this finding. One study found that the DLPFC was activated during temporary maintenance of the position of sequentially presented spatial stimuli, while another study found that activation of the DLPFC was sustained during the retention of spatial stimuli when capacity demands were high. A plausible alternate explanation for these findings is that the forward condition of the Spatial Span Test may have been a better measure of complex executive attention than the backward condition of the test.

Assuming that a practice effect occurred from the forward to the backward condition, the executive element may have been removed through practice, leaving the forward condition as the better measure of executive attention.

The mean number of digits recalled on the forward condition of the Digit Span Test (6.77) was higher than the mean amount of items recalled in the correct sequence on the forward condition of the Spatial Span Test (5.84). Baddeley (1996c, 2003) confirmed that there was usually a two-point discrepancy between the Digit Span and Corsi Span tasks in neurologically intact subjects. Lezak (2004) noted that there was a minimum difference of one item between the Digit Span and Corsi Span tasks. This finding was well-documented and established, but had not been explored in detail in the literature. Greater demand placed on the central executive on the Spatial Span Test would theoretically result in fewer executive resources for performance, and could account for the lower spatial than digit span mean scores.

Lezak (2004) contended that should two tests be so alike that they could be treated as equivalent tests that measured exactly the same cognitive abilities, inclusion of both tests would be unnecessary in a test battery. The backward condition of the Spatial Span Test seems to have been a fairly redundant test that was complicated by a practice effect. Only the forward condition of the test may be necessary for scoring purposes (Lezak, 2004; Wilde et al., 2004). As the backward condition of the Spatial Span Test did not correlate significantly with any other test it was subsequently removed from further analysis, and the correlation matrix was run a second time without this variable.

Complex Executive Attention

The study hypothesised that the *n*- Back Test and the backward conditions of the Digit Span and Spatial Span Tests would correlate with part B of the Trail Making Test, and with the interference and mixed interference phases of the Stroop Test according to the use of domain-free complex executive attention. The correlation matrix revealed that only the 2-Back condition of the *n*- Back Test correlated significantly with a number of variables. The 1-Back and 3-Back conditions of the *n*-

Back Test each correlated significantly with the 2-Back condition of the test, but did not correlate significantly with each other, or with any other variable (Table 6). The 2-Back condition of the *n*- Back Test correlated moderately and negatively, but significantly, with performance on Part B of the Trail Making Test (r = - .30, p < .015), as well as moderately and significantly with the backward condition of the Digit Span Test (r = .31, p < .015), as predicted in the hypotheses.

The fact that the 3- Back condition of the *n* – Back Test only correlated with the 2-Back condition, and did not correlate significantly with any other variable supported research which suggested that the 3-Back condition may not be a valid measure of working memory (Owen et al., 2005). This study included a 3-back condition in order to test the limits of performance on this task. The 3-back was not excluded from the analysis because it correlated moderately, but significantly with the 2-Back subtest (r = 0.27, p < .015). However, only 52 of the 105 participants were able to complete the 3-Back condition, and most of the scores on this subtest were confounded by high commission error.

Accuracy on the *n*-Back Test was used as the key measure of working memory performance on this task. However, the concept of accuracy merits consideration. First, accuracy did not necessarily reflect true performance on this task. Performance was also affected by and reflected in the number of omission and commission errors the test-taker produced. Omission error in this case refers to the number of target items that are not responded to, which is a reflection of true error whereby the subject is unaware that an item is a target. Commission error refers to the number of false alarm errors whereby the test-taker either randomly responds to an incorrect target, or responds to "lure" or "foil" targets. Kane et al. (2007) asserted that a 'lure effect' refers to items in a sequence that the test-taker incorrectly responds towards because the item is familiar as it appeared previously in the sequence, or it resembles the target item. In this study, the number of errors – both omission and commission – on the *n*- Back Test increased significantly as *n* increased (Appendix I). This finding supports research that suggests that performance on the 3-Back subtest may not be a valid measure of working memory (Owen et al., 2005)

Contrary to expectation, the 2-Back condition of the *n*- Back Test correlated weakly but significantly with the forward condition of the Spatial Span Test (r = .248, p < .015). Under the assumption that the forward condition of the Spatial Span Test tapped into complex executive attention due to the cognitive demands elicited by serial sequential processing of spatial information, the significant correlation of the forward condition of the Spatial Span test with the 2-Back condition of the *n*-Back Test could be explained. The correlation could reflect a relationship between these tests based on the utilisation of domain-free complex executive attention, as predicted in the hypotheses of the study. Also contrary to expectation, the *n*- Back Test did not correlate significantly with the interference and mixed interference phases of the Stroop Test. The Stroop Test only correlated with performance on the Trail Making Test, and with performance on the Computer Mouse Test (Table 6). Even after computer mouse performance had been partialed from the analysis, the four phases of the Stroop Test remained highly inter-correlated. The strong relationships within the Stroop Test were not consistent with any of the study hypotheses.

A possible explanation for why the interference and mixed interference phases of the Stroop Test did not correlate with performance on the backward condition of the Digit Span Test or with the *n*- Back Test is that the computer-based version of the Stroop Test may not have produced a significant interference effect, due to the fact that the subjects were not required to articulate the words. Gazzaniga et al. (2002) proposed that articulation of the words on the Stroop task was necessary for verbal interference. It had been demonstrated that verbal presentation of material had a more pronounced effect than visual presentation of verbal material in working memory (Logie, 1995). Visually presented information had to first be encoded into a speech-based format before entering the phonological store. The interference between acoustically presented items and items that were held in the phonological store, no matter how briefly stored, produced the typical interference effect. Although the Stroop Test was not a working memory test, it demanded the same control over focused attention and the ability to ignore verbal interference. The delay between visual presentation of the information and encoding may have given subjects enough time to dismiss the effects of interference on the Stroop task and give the correct answer. The low incidence of error on the Stroop test (Appendix I) could provide evidence to support this view.

Each subtest of the Stroop Test correlated significantly with both parts of the Trail Making Test (Table 6). The relationships between the Stroop and the Trail Making Tests revealed fairly strong correlations that indicated that these tests did not necessarily measure the same construct, but measured similar constructs. Both the Stroop and Trail Making Tests also correlated robustly with the computer mouse test. It is probable that the Stroop and Trail Making Tests correlated within and between each other around the common component of speed. The relationships between the speeded Stroop, Trail Making, and computer mouse Tests were consistent with the hypothesis that computer ability would affect performance on the speeded computer-based tests.

Part A of the computer-based Trail Making Test correlated fairly robustly with part B of the test (r = .56, p <.0001). Previous research had also found that Part B correlated significantly with Part A of the test (Lezak, 2004). The relationship indicated that these subtests measured a similar cognitive construct. However, the correlation was not high enough to suggest that these subtests measured the same construct. One explanation for this finding is that both parts of the Trail Making Test may have correlated with each other around the common component of speed. An alternate explanation is that there may have been a practise effect from part A to part B, so that better control of the computer mouse facilitated performance on part B of the test.

The forward and backward conditions of the Digit Span Test were also significantly correlated (r = .42, p < .0001). The correlation was also not high enough to suggest that these subtests measured the same cognitive construct. The raw data demonstrated that performance on the backward phase of the Digit Span Test was better than on the forward phase in 12 subjects, or approximately 11% of the sample. There was no difference in performance between the forward and backward conditions of the Digit Span Test in 18 subjects, or approximately 17% of the sample. Altogether, 28% of the sample performed at least as well on the backward as on the forward phase of the test.

Analysis of the raw scores suggested that backward span was perhaps only a little more challenging than forward span in the study.

One explanation for this finding is that the computer-based Digit Span Test differed from the traditional version of the test as it allowed test-takers to view the sequence of numbers that they entered on the computer screen, so that they could alter errors in the sequence before moving to the next task. On the backward phase of the task, allowing the test-takers to view the sequence on the computer screen may have prompted them to read the numbers from last to first, thereby transforming the backward task into a similar measure to the forward task. Lezak (2004) cited research that found that mental imagery played a role in backward span performance, as test-takers created an image of the numbers and then “read” them backwards. Patients with spatial inattention could not perform this task effectively, providing evidence for the function of mental imagery in backward span performance (Lezak, 2004). Some of the executive components of the backward condition could theoretically be removed if the test-taker did not have to create a mental image of, or manipulate the numbers in, memory. An alternate explanation is that the computer-based Digit Span Test allowed the subject an unlimited amount of time to answer before moving onto the next sequence. It is possible that some test-takers rehearsed the information so that it transferred to long-term memory, and that performance on the backward condition was facilitated by repetition. Conway et al. (2005) contended that this kind of issue could arise in self-paced tasks where there were long delays between presentations of stimuli.

Processing of an additional task while maintaining information in memory constitutes the executive component of working memory that theoretically transforms the backward condition of the Digit Span Test into a working memory measure (Conklin et al., 2000; Conway et al., 2005). Removal of the central executive element on the backward condition of the test could explain the relatively large correlation between the forward and backward conditions of the test. On the other hand, the correlation was not large enough to suggest that these tests measured the same construct. Only 28% of the sample performed at least as well on the backward as the forward condition. This suggests that for 72% of the sample, the backward condition of the

Digit Span Test was still more demanding, and utilised more executive resources, than the forward condition of the test.

In summary, the 2- Back condition of the *n*- Back Test, Part B of the Trail Making Test and the backward condition of the Digits Span Tests appear to have correlated with one another around the use of complex executive attention, as predicted in the study hypotheses. The significant correlation of the forward condition of the Spatial Span Test with the 2-Back condition of the *n*- Back Test was not expected, but could be explained by the central executive demands of Spatial Span performance. None of the correlations was high enough to suggest that the tests measured the same processes, and so examination of the factor structure of the correlation matrix was performed to see whether the correlations between the tests could be explained by a few factors.

4.2.3 The validity of the *n*- Back Test

Exploratory factor analysis was selected as the best statistical method for examining the latent relationships between the variables of the study. Factor analysis reveals which variables collectively measure the same underlying cognitive constructs (Conway, Kane, & Engle, 2003). The primary aim of the factor analysis in this study was to investigate the construct validity of the *n*-Back Test (Kane et al., 2007). Construct validity examines how well a test measures a specific construct (Gibertini & Retzlaff, 1994). It is evaluated by means of a correlation matrix, and examines the degree to which a construct that is measured by different tests is sustained across the tests. Convergent and divergent validity may be found in the correlation matrix, as the tests that measure the same cognitive ability would, theoretically, "…converge on one another and diverge from the rest of the pack" (Conway et al., 2005, p. 781).

Under the assumption that the *n*-Back Test measured working memory capacity as processing load increased, it was hypothesised that the *n*- Back Test would converge with the backward conditions of the Digit Span and Spatial Span Tests as measures of working memory span, and would diverge from the forward conditions of the Digit Span and Spatial Span Tests as measures of simple short-term memory span. In other words for the *n*-Back Test to demonstrate construct validity as a measure of working memory, the test had to load together with tests that manipulated information in working memory rather than with those that relied primarily on simple rehearsal or maintenance of information in short-term memory (Kane et al., 2007).

Limited research had compared the *n*- Back task with other established measures of working memory such as reading span and operation span tasks (Kane et al., 2007). Some evidence against the *n*-Back Test as an adequate measure of working memory had been found. One study that compared ‘complex’ span with the *n*-Back task suggested that the *n*-Back Test was better correlated with simple span tasks than with the more complicated ‘complex’ span task, implying that there was little or no association between the *n*-Back Test and working memory span (Kane et al., 2007).

However, this finding was problematic for a number of reasons. First, the researchers had created a composite score for a 2- to 5- Back *n*- Back Test. A composite score for the *n*- Back Test is methodologically problematic, as each phase of the *n*-Back Test taps into different aspects of working memory. For example, a 2-Back task is significantly more difficult than a 1-Back task and thus cannot be assumed to measure the same processes as a 1-Back task. A composite score therefore may not adequately reflect the processes differentially measured by each phase of the test. Second, the *n*-Back Test used in the study ranged from a two-back to a five-back task. This method may be questioned, as most *n*- Back tests vary the load up to 3-back and, even here, the validity of the task has been questioned as the ability to perform the test declines (Owen et al., 2005).

Carpenter et al. (2000) maintained that neural activity in the DLPFC increased up to a 2-back task then decreased on a 3-back task in neuroimaging studies, reflecting a behavioural and neurological drop in performance after the 2-back condition of the test. If the validity of the 3-Back subtest has been questioned, then the validity of a 4- or 5- Back task should be questioned too.

The third concern is that ‘complex’ span was not adequately defined in the study, and has not been adequately defined in the literature (Baddeley, 2000b). Working memory span measures, such as reading span, operation span and sentence span tasks involve complicated relationships between a number of cognitive systems in the brain. These relationships are significant, and complex span predicts performance on important cognitive constructs, such as intelligence (Kane et al., 2007). However, the constructs measured by working memory span tasks are not explicit, and more research needs to be conducted on these tests to examine precisely what aspect of cognitive functioning they measure (Baddeley, 2000b).

Some evidence for the *n*- Back Test as an adequate measure of working memory has been found. One study found a relationship between faster responses on the 2-back condition of the *n*-Back Test and increased intelligence (IQ), which pointed to a relationship between intelligence and *n*-Back performance (Kane et al., 2007). This relationship was important, as the test had been criticised for not being able to predict performance on measures of intelligence. Performance on the *n*-Back Test has been linked with performance on measures of attentional control (Kane et al., 2007). Cognitive performance on the *n*- Back Test has also been mirrored by neurophysiological response in the DLPFC and other areas that tap into the executive functions of the brain (Gazzaniga et al., 2002).

Initial Findings

The results of this study revealed fairly simple structure on an orthogonal rotation of the factor structure. Simple structure is crucial to factor analysis as it makes the factor pattern replicable and easy to interpret (Kline, 1994). The first criterion for simple structure was that every row of the rotated factor structure had to include a loading close to zero. The second criterion was that within each factor, there had to be at least as many zero loadings as there were factors. The third criterion was that each variable should load significantly onto one factor but should have zero loadings on other factors. The fourth criterion was that there should be proportionately more zero loadings than significant loadings on a factor. The final criteria were that the items loadings on each factor had to be greater than 0.30 to be significant and that items should ideally load significantly onto only one factor (Kline, 1994). Orthogonal rotation of the factor analysis revealed fairly simple structure, according to the criteria outline above. In addition, examination of the residual matrix demonstrated low correlations (Appendix G). The residual matrix provided additional evidence that simple structure was attained for the factor analysis.

Three factors were revealed in an orthogonally rotated factor analysis. The first factor comprised significant loadings on Stroop1, Stroop2, Stroop3, Stroop 4, and parts A and B of the Trail Making Test (Table 10). The second factor comprised significant loadings on the 1- 2- and 3-Back conditions of the *n*-Back Test. The third factor comprised significant loadings on the forward and backward phases of the Digit Span Test. Surprisingly; the forward condition of the Spatial Span Test did not load significantly onto any factor, but loaded equally and partially at .33 on both the second and third factors.

The first factor grouped together the Stroop Test and the Trail Making Tests. Both of these tests had no memory component but placed heavy demands on executive focusing and switching of attention in the latter phases of the tests. The Stroop and Trail Making Tests were also the only time-based or speeded tests utilised in the study. The fact that the word-reading and colour-naming subtests of the Stroop Test loaded significantly onto the same factor as the interference and mixed interference subtests of the Stroop Test meant that the subtests were not differentiated in terms of the demands placed on executive attention, but were grouped together as tests that relied on processing speed. This concept was further substantiated by the fact that part A and part B of the Trail Making Test also loaded significantly onto factor 1, were not differentiated in terms of the demand placed on executive attention, relied on processing speed, and were grouped together with the Stroop Test.

Based on the grouping of the variables, ‘processing speed’ was initially selected as the most comprehensible label for the first factor. However, although computer mouse ability had been partialed from the analysis, there was almost certainly a third variable effect of computer familiarity on performance on these tests, as demonstrated in the *t*-tests and in the significant correlations between computer mouse performance and performance on the Stroop and Trail Making Tests. Removal of computer mouse test scores may have removed most of the variance caused by these factors, but it is possible that other variables associated with computer familiarity continued to affect performance, particularly on the speeded tests. This may have prevented the Stroop and Trail Making Tests from being accurately represented in the factor analysis. The first factor was subsequently re-labelled ‘processing speed, informed by computer familiarity’. An alternate explanation is that the Stroop and Trail Making Tests measured the same executive process, and that they factored together as measures of this process without a memory component.

The study hypothesised that the *n*-Back Test would correlate and factor together with the backward condition of the Digit Span Test and the backward condition of the Spatial Span Test as measures of working memory span. The results of the analysis did not support this hypothesis. The study also hypothesised that the forward condition of the Digit Span Test and the forward condition of the Spatial Span Test would correlate as measures of simple memory span. However, the results once again did not support the hypotheses of the study. To explain these findings, the basic cognitive requirements and processing abilities demanded by each task are reconsidered below.

Performance on the *n*-Back Test required constant monitoring of a set of serially presented items, and identification of items that appeared *n* items previously in the sequence (Owen et al., 2005). The task required subjects to continuously update new information, whilst simultaneously recalling the last few items in a sequence, and to selectively forget previous items when they were no longer relevant to the task. The *n*-Back Test required both maintenance of information in short-term memory, and simultaneous processing of an additional task (incoming information), thereby meeting the theoretical requirements for working memory performance (Conway et al., 2005). The task also relied on the ability to *switch* attention rapidly between tasks that required constant monitoring of new information, whilst updating old information, which was a separable function of the central executive component of working memory (Baddeley, 1996a)

The *n*-Back Test used in this study was a visual task, and was designed in such a way as to actively prevent the items from being recoded to verbal format. The *n*- Back

Test theoretically accessed and utilised both central executive processes and the visuospatial sketchpad for performance, particularly on the 2-Back condition of the test. Baddeley (1996c) maintained that encoding, manipulation, and retrieval of visuospatial information relied chiefly on the central executive component of working memory. The visual imagery and spatial sub-systems of working memory were not as easy to encode as the phonological system and the sketchpad was subsequently more demanding on the central executive (Logie, 1995). Additionally, due to the inter-stimulus time limit imposed by the test, there was not enough time for subjects to transfer items from the *n*-Back Test to long-term memory.

The forward condition of the Spatial Span Test demanded monitoring of a set of serially and sequentially presented spatial items, and immediate recall of only the order of the items (Wilde et al., 2004). It did not theoretically meet the requirements for working memory capacity, as it relied primarily on short-term maintenance of information (De Lillo, 2004). However, Rudkin et al. (2007) proposed that processing of serial sequential spatial information placed greater demands on the central executive than serial processing of verbal information. Spatial information was also not easily encoded to long-term memory, and so was theoretically more demanding on the central executive than the phonological loop (Logie, 1995). Common cognitive requirements and processing abilities required for performance on both the Spatial Span Test and the 2-Back condition of the *n*- Back Test included serial sequential processing, short-term retention of visuospatial information, and use of complex executive attention for performance.

The Digit Span Test required monitoring of a set of serially presented numbers, and recall of both the individual numbers and in the order in which they were presented. The forward condition of the test did not theoretically meet the requirements for working memory capacity, as the task relied primarily on short-term maintenance of information. The backward condition of the test theoretically relied on both maintenance of information and simultaneous processing of an additional task (manipulating the digits in memory), thereby meeting the requirements for working memory performance. However, computer-based administration of the Digit Span Test may have undermined the validity of the backward condition of the test, making it an easier task than the traditional version of the test. Additionally, the unlimited time given to test-takers to respond to each sequence may have assisted with the transfer of verbal material to long-term memory on the Digit Span Test. The conditions for working memory span on the backward condition of the Digit Span Test in this study may not have been entirely met. The Digit Span Test theoretically relied on the episodic buffer component of the central executive in order to *activate, maintain* and *manipulate* information from *long-term memory* , a separable function of the central executive (Baddeley, 1996a)

Rudkin et al. (2007) proposed that serial recall of verbal information was less demanding on the central executive than serial sequential recall of spatial information. One of the reasons for this proposed difference is that the Digit Span Test accessed and utilised the phonological loop for rehearsal of information in working memory. The ability to rehearse information made the verbal component of the working memory system less demanding on the central executive than the visuospatial sketchpad, which did not have the same rehearsal abilities and access to pre-existing information in long-term memory as the phonological loop (Logie, 1995).

One explanation for the results of the factor analysis is that the *n*-Back Test was a more difficult test than the Digit Span Test, relying on more extensive use of the central executive component of working memory This explanation would support the concept that each of these tests measured a single construct with each task revealing different points on a continuum (Conway et al., 2005). However, if this were the case the tests would theoretically have loaded onto a single factor, with stronger or weaker item loadings reflecting which subtests were better or worse measures of the construct.

An alternate explanation is that the n-Back and Digit Span Tests loaded onto different factors according to the separable function of the central executive component of working memory measured by each test, and to whether the test met the requirements for measuring working memory capacity or not. The second factor was comprised of significant loadings on the 1-Back (.79), 2- Back (.78), and 3-Back (.62) conditions of the *n*-Back Test. The 1- and 2- Back subtests loaded more significantly on this factor than the 3-Back subtest, which suggests that they may be better measures of the construct measured by the factor than the 3-Back test. The subtests of the *n*-Back Test may have factored together around the executive ability to *switch* attention rapidly between tasks that required constant monitoring in working memory. The *n*-back task also met the theoretical requirements for working memory span.

The third factor was comprised of significant loadings on the forward (.83) and backward (.74) subtests of the Digit Span Test. The forward and backward conditions of the Digit Span Test may have factored together around short-term memory span and the less extensive involvement of the central executive, due to the relative autonomy of the phonological loop in maintaining verbal information in short-term memory. The relatively higher loading of the forward condition on this factor may reflect the fact that this subtest was a better measure of short-term memory than the backward condition, which required a few more executive resources for performance than the forward condition of the test. These tasks may also have factored together around the separable executive ability to *activate*, *maintain* and *manipulate* information from long-term memory.

The forward condition of the Spatial Span Test did not load significantly on either factor, which suggests that it did not measure the same construct as either of the factors, but that it did share some processes with both constructs measured by the *n*-Back and Digit Span Tests. Due to the fact that the Spatial Span Test was administered in exactly the same manner as the Digit Span Test, some of the variance on the Spatial Span Test may be explained as a measure of serial sequential maintenance of items in memory. The rest of the variance on the Spatial Span Test could be explained as a measure of complex executive attention, and the more demanding executive resources needed for serial sequential spatial performance than for serial verbal performance, as reflected in the partial loading of the Spatial Span Test on the same factor as the *n*-Back Test.

An alternate explanation is that the subtests of the *n*-Back Test may have factored together as measures of visual working memory, whereas the forward and backward conditions of the Digit Span Test factored together as measures of verbal working memory, and that the tests were not distinguished in terms of their relative demand on central executive processes but merely reflected the domain-specific components of working memory measured by each test. However, this explanation would not be able to account for the partial, and equal, loading of the Spatial Span Test onto both the second and third factors.

The results suggest that the question of the validity of the *n*- Back Test could not be answered sufficiently from the results of the factor analysis. However, the correlation matrix may provide a better case for the validity of the *n*-Back Test. First, the 2-Back condition of the *n*- Back Test converged with other tests that were assumed to tap into central executive processes. It shared a significant relationship with part B of the Trail Making Test (r = -30, p < .01), with the backward condition of the Digit Span Test (r = .31, p < .01), and with the forward condition of the Spatial Span Test (r = .248, p < .01). Second, the 2-Back condition of the *n*-Back Test diverged from tests that theoretically measured simple maintenance of information in short-term memory, such as the forward condition of the Digit Span Test (r = .15, n.s.). The 2-Back subtest of the *n*-Back Test does in this case seem to demonstrate some construct validity as a measure of the executive component of working memory.

4.3 Limitations of the Study

One of the limitations of the study was that performance on the tests of working memory was statistically aggregated. In other words, the study did not account for individual differences in performance on the tests. Logie (1995) maintained that individual patterns of performance on cognitive-behavioural tests is seldom detailed in the literature, and may be integral to understanding unique strategic techniques within working memory task performance. For example, the kind of error produced by the test-taker contains a lot of important information, including the type of judgement-call made by the subject. Some subjects respond impulsively to almost all targets, which results in seemingly more accurate response to targets but exceedingly high commission error. In a clinical context, impulsivity may be a marker for poor executive control or frontal lobe compromise. Individual differences are highly important and need to be investigated for performance on the tests to have relevance in a clinical context.

Reliability for the tests could not be assessed due to the nature of the tests, and due to the time limits imposed by the study. Ascertaining the reliability of the tests is important for future development and use of the computer-based versions of the tests used in this study and it is both feasible and necessary to investigate the reliabilities of the tests in future research. To assess test-retest reliability, there would need to be a large time gap between administrations of the tests and it would be necessary to assess whether practice effects have occurred between administrations, as practice effects have been consistently demonstrated on many of the tests.

The development of alternate forms in future research may also be a valuable tool for investigating the internal consistency of the tests. Practise effects often occur on repeated administration of the same measure. The creation of alternate forms may be useful in a clinical context where a patient’s recovery is tracked over a period of time, so as to undermine practice effects and track genuine change in cognitive functioning. With the implementation of a cut-off point on the speeded tests, it may be possible in future research to investigate the internal consistency of the Stroop and Trail Making Tests. Future development of the *n*-Back Test could also ensure that the items relate to one another so that internal consistency on each of the phases of the test may be assessed.

Another limitation with the study was that there were potential threats to the validity of the computerised Stroop and Digit Span Tests. In particular, computerised administration of the Digit Span Test may have transformed the backward condition of the test to a similar measure to the forward condition of the test through visual presentation of the digits on the computer screen and due to lack of a time limit for response. Though it was clear that the forward and backward conditions of the computer-based Digit Span Test did not measure the same construct, much of the executive element may have been removed on the backward condition of the test. Computerised administration of the Stroop Test may also have threatened the validity of this test. Visual presentation of the words seems to have undermined the interference effect found on the original test (Gazzaniga et al., 2002).

The sample for the study was very specific and so the generalisability of the findings may be limited. The sample was a well-educated, adult and volunteer sample. The sample was screened for diagnosed neurological, psychiatric or learning disabilities. The results of this study may not be generalised to children whose working memory system is underdeveloped; to elderly people whose working memory system may have begun to fail; to a clinical neurological population whose working memory system may be organically damaged; to a psychiatric or learning disabled population; or to a less educated population whose access to information technology may be severely compromised and who are not test-wise. Dichotomising the variables of the study may also have resulted in the loss of detailed information about differences in performance between different groups.

However, the study raised some important issues for the South African context, most important of which was the finding that even highly educated South African adults were affected by familiarity and experience using a computer on the tests. This finding needs to be further explored in different groups, in order to begin to understand potential sources of bias in computer-based testing in South Africa. This research could be conducted in a positive way to find a way to account for these sources of bias so that computer based testing can be used in a fair manner, taking context into account.

4.4 Implications of the Study for Theory and Research

Although the sample for the study was fairly narrowly defined, the results of this study could provide a useful baseline of performance against which test scores may be compared, particularly in a well-educated population. The impact of working memory deficit can be devastating, particularly in the workplace and in jobs that demand high levels of working memory capacity and central executive resources. People with high pre-morbid levels of education tend to perform better than people with lower levels of education post-injury or disease (Howieson et al., 2004). One of the current challenges in neuropsychological assessment is to be able to detect working memory impairment, utilising tests that can detect even subtle impairment in highly educated individuals. The results of this study may begin to provide an idea of what kind of scores can be expected in a well-educated sample on the tests utilised in this research.

For example, the 0-Back condition of the *n*- Back Test was excluded from the analysis, but may be a useful baseline measure of performance against which performance on the 1-Back, 2-Back and 3-Back conditions could be compared. The fact that the majority of the sample performed the 0-Back condition 100% correctly suggests that adults with at least 12 years of formal education should be able to perform the 0-Back Test accurately. As such, performance on the 0-Back condition may have practical implications as poor performance on this task may indicate an inability to sustain attention in a clinical population, or it could indicate malingering in the testing context.

Another theoretical implication of the findings is that the 3-Back condition of the *n*-Back Test may not be a valid measure of working memory, as even well-educated subjects struggled to complete the task successfully. However, the test used in this study was visual, and may have inadvertently placed greater demand on the central executive than a verbal version of the test. The theoretical implication of this finding for future research is that verbal 3-Back performance should be compared with visual and spatial 3-Back performance to see whether performance on the 3-Back condition is perhaps more valid when the *n*-Back Test is presented in different domains.

The results of the study suggest that the psychometric properties of computer-based tests should be further explored in the South African context, particularly as performance on a computer mouse test has been shown to create statistical differences between groups who, due to social, cultural and historical factors, may have had different access and exposure to computers and other information technology tools. The findings of this study highlight that neuropsychological test results cannot be interpreted without carefully examining the social, cultural and historical context in which test scores are acquired (Grieve, 2005). However, there are some significant advantages to computer-based test administration. The benefits of computer-based tests may outweigh the adverse effects of their use, provided that the variables that may influence test performance in the South African context are accounted for (Bush et al., 2002). Some of the advantages and disadvantages of the computer-based tests utilised in this study, and practical implications of the study, are outlined below.

Computer-based administration of the Digit Span Test improved upon the traditional test, as the presentation of instructions and administration of the test was standardised, the rate at which the numbers were articulated was accurate and on time, the clarity of articulation was constant, and subject's responses were accurately recorded to enhance the standardisation of the test (Davies et al., 2005). However, the computer-based Digit Span Test differed in two ways from the traditional test. First, the subject was required to enter the digits into the computer keyboard as opposed to articulating the digits out loud to the examiner. Second, the test-taker's response was visually presented on the computer screen so that errors could be corrected before moving on to the next sequence of the test. Practical implications of the computerised administration of the test include the fact that computer-based administration may have undermined the validity of this test. These findings imply that the Digit Span Test used in this study may not be appropriate for use in a clinical context at present, and that changes need to be made to combat threats to the validity of this test.

A number of studies proposed that computerised administration of the Spatial Span Test was more useful than traditional administration of the test, and that differences in presentation did not affect performance on this test (De Lillo, 2004). De Lillo (2004) asserted that computer-based presentation of spatial stimuli was advantageous, as the stimuli could be represented in both two and three dimensions, vertically and horizontally, and the spatial path could be represented by an increase in brightness or a change in the colour of the stimulus. Computer-based recording of response also provides information about the amount of time that the subject takes to respond; a variable that has been shown to reveal significant characteristics of how subjects plan and strategise (De Lillo, 2004). The findings from this study suggest that the backward span of the Spatial Span Test may need to be examined in relation to its validity as a measure of working memory capacity. However, the author did not find any administrative problems with the study that could account for the finding that 47% of the sample performed at least as well on the forward as the backward condition of the test. The findings imply that there may be an issue with using the backward Spatial Span Test as a valid measure of working memory capacity in the clinical context.

Computer-based administration of the Stroop and Trail Making Tests improved upon traditional ‘paper and pencil’ administration in a number of ways. First, it improved the accuracy of timing the response of the test-taker. Second, the computer-based tests did not allow the test-taker to move onto the next item until the current item was correctly responded towards, which controlled for the variability in administration that occurred when an examiner failed to identify a mistake. Both speed and error could be accurately recorded on the computer-based version of the tests. However, the computerised versions could not identify whether the test-taker perseverated, was impulsive, or whether there were motor difficulties that prevented the subject from performing the tests at speed. These factors need to be accounted for, particularly in clinical populations where impulsivity and motor difficulty may indicate brain dysfunction.

Computer-based administration of the Trail Making Test differed from the traditional administration, as the test-taker was required to click on the stimuli via the computer mouse. It subsequently relied on the ability to operate a computer mouse effectively and at speed. The results of the study demonstrated that speed on a computer mouse was significantly related to performance on this test. The results of the factor analysis suggested that the Trail Making and Stroop Tests factored together around the common construct of processing speed. The results of the study imply that a baseline measure of performance on a computer mouse test needs to be taken and partialed from performance on the Trail Making Test.

Computerised administration of the Stroop Test also differed from traditional administration, as each stimulus was presented singly in the centre of the computer screen. The test-taker was required to click on one of four buttons to indicate the correct answer for a particular stimulus before moving on to the next stimulus. The time scores recorded on these tests could not be directly compared with time scores on the traditional version of the test, as computer-based administration slowed the response of the test-taker on the test. The results of the study suggested that visual presentation of the stimuli and computer-based response did not elicit an interference effect on this test (Gazzaniga et al, 2002). The results of the study imply that computerised administration of this test may not be beneficial, and that the validity of the test may be undermined through computer-based presentation of test stimuli.

The *n*-Back Test was originally a computer-based test. The version of the test used in this study was computer-paced, and the objective of computer-pacing the task was to control and standardise the administration of the test. Following the work of Barrouillet et al (2007), standardising the time for presentation of stimuli and for the inter-stimulus delay was implemented to restrict the ability to rehearse the information, not by preventing the subject from switching attention but to examine the test-taker's ability to switch attention within a controlled time frame. This step was implemented in the study to prevent recoding of information to long-term memory. The current version of the n- Back Test lengthened the inter-stimulus interval by 500ms per phase, to accommodate the increase in cognitive demand on each phase of the test. The results of the study suggested that the 1- and 2-Back condition of the test were useful measures of the executive ability to switch attention rapidly between tasks under a time constraint. The results suggested that the 2-Back condition was a good measure of complex executive attention, reflected in its strong correlations with part B of the Trail Making Test, the backward condition of the Digit Span Test and the forward condition of the Spatial Span Test. Theoretical implications of this finding include that the 2-Back condition of the n-Back Test may be a valid measure of working memory capacity, and that further investigation of this test is necessary to determine its validity.

The study revealed that the validity of test scores on the 3-Back condition of the test could be confounded by high commission error. One of the implications of this finding is that a thorough theoretical investigation into the effects of presentation time and inter-stimulus delay is necessary for the development and use of the *n*-Back Test in future research (Hancock et al., 2007). Practical implications of the study include measuring both omission and commission error, and statistically correcting for the effect of this variable on test performance. In effect, future development of the test should account for the effect of "false positives" in terms of scoring performance on the test.

The results of the study suggest that the tests correlated within each other around the executive component measured by each test, and also perhaps by the verbal, spatial or visual domain measured by each test. The results suggest that the backward condition of the Digit Span Test correlated with the 2-Back condition of the *n*-Back Test around the use of executive attention, and it correlated with the forward condition of the Digit Span Test as a serial sequential verbal memory measure. The Digit Span Test did not correlate with any other test, which could imply that it was a domain-specific verbal test. The results suggest that Part B of the Trail Making Test may have correlated with the 2-Back condition of the *n*-Back Test around the use of executive attention, and may have correlated with Part A of the Trail Making Test as a memory-free measure of attention. The forward condition of the Spatial Span Test may have correlated with the 2-Back condition of the *n*-Back Test as a measure of executive attention. It did not correlate with any other test, which may imply that as the test was the only spatial test, it remained uncorrelated with tests that measured other domains of working memory.

Conway et al. (2005) argued that, when examining a cognitive construct, it was "…best to use multiple, reliable measures that do not replicate one another" (p.780). The results of the study suggest that working memory may be assessed using a variety of tasks that each illuminates an aspect of working memory performance. Baddeley (2000b) contended that research should turn towards identifying both the central executive processes of working memory, and the individual roles that the phonological loop and visuo-spatial sketchpad play in working memory performance. For example, in Alzheimer's patients, visuo-spatial and phonological working memory may individually be relatively unimpaired, but the ability to combine performance on the subsystems, a central executive function, is often impaired (Baddeley, 2000b). It may be more useful in a clinical context to measure and assess each aspect of working memory individually, in order to fully comprehend the nature of working memory, and to determine where a deficit may lie. Future research should perhaps investigate the benefit of using a combination of tests to measure the construct of working memory relative to using only one “working memory span” test to measure this construct.

4.5 Recommendations and directions for Future Research

In South Africa, psychometric tests have not always been contextually sensitive (Foxcroft, 1997). In the current South African context, the use of computer-based tests must be carefully examined. A particular concern is that a large portion of the population may not have adequate access to computers, and may not be adequately trained in or exposed to computer technology (Davies et al., 2005). Future research on the use of computer-based tests should examine the psychometric properties of such tests particularly amongst sectors of the population that may not have been adequately exposed to information technology tools. A positive finding from the results of this study is that the validity of computer-based tests may be enhanced if a baseline of performance is taken before the test-taker begins to perform the tests, and if that baseline is taken into account when interpreting results, particularly on speeded neuropsychological tests. The practical implications of the findings include making a few important changes to the tests to enhance their validity. A recommendation from this study is that these changes should be implemented and that the validity of the tests should be investigated thoroughly in future research.

Future development of the Digit Span Test should include standardising the inter-item response time on both the forward and backward condition of the task, not permitting test-takers to view their response, and voice-recording the subject’s response. These changes could enhance the validity of the computer-based administration of the test. Future development of the Stroop Test should include computer-based presentation of stimuli and voice-recording the subject’s response, or timing the test and allowing the examiner to press a key to indicate when the subject has made an error. These changes may enhance the validity of these tests.

In terms of the assessment of working memory, an interesting and important finding from fMRI research is that areas of the brain integrated in different ways according to the requirements of a task and according to individual cognitive strategies and skills that were developed within a particular culture (Carpenter et al., 2000). This concept is important for future assessment of working memory, as it suggests that neural networks are configured over time and may be organised and informed by cultural learning. Logie (1995) maintained that the pattern of performance of individuals on cognitive-behavioural tests was seldom detailed in the literature, and may be integral to understanding distinctive strategic techniques within working memory task performance. A recommendation from this study is that individual differences in performance on these tests should be investigated, particularly with regard to culturally-based differences in performance.

A recommendation for improving the physical administration of the tests is put forward. Hausler, Sommer and Chroust (2007) proposed a number of suggestions for minimising the problem of technical errors of measurement in computer-based tests that are run on Windows platforms. These suggestions are particularly pertinent to the development of the tests utilised in the study and, according to international guidelines on computer-based and internet-delivered assessment, should be taken into account in the future development of computer-based tests (ITC guidelines, 2005). The amount of hardware that interferes between a subject's response and the computer's recording of the response may cause problems with the timing of the response. This can in turn cause variability in the recording of the response that is not due to variability in the subject's response (Gur et al., 2001). The following suggestions refer mainly to computer-based tests that include some measure of timing.

The first suggestion for minimising technical errors of measurement in computer-based tests is to close programs that are running in the background during the administration of a test, as these programs can interrupt recording of responses and the timing of item presentation (Hausler et al., 2007). The second suggestion for minimising errors of measurement is to account for the rate of presentation of stimuli and to be able to predict how long it would take for a particular graphic to appear on the computer screen. The measurement errors that arise from this kind of variability in presentation include inadequate time to respond due to delays in the presentation of graphics, and to reduced recording of response, due to the fact that graphics presentation delays may take up the entire duration of the inter-stimulus response interval in timed tests (Hausler et al., 2007). The third suggestion for minimising errors of measurement is to investigate the amount of time that it takes for the detection of the reaction of the subject via a particular recording device such as a touch screen or computer keyboard. The time that it takes to register a response should be accounted for in the timing of the test, and in the amount of time given to the subject to respond to the stimulus.

Another direction for future research and development of computer-based tests would be to move towards touch screen computer-based assessment. The Brain Resource Company (BRC) of Australia has developed a touch-screen-based set of neurocognitive tests that includes many of the tests utilised in this study. They report high reliability and validity between their computer-based versions and the original versions of the tests (Paul et al., 2005). Computerised Adaptive Testing (CAT) is another method of computer-based assessment that may be particularly useful in the South African context. CAT ensures that if a test begins at a level that is too difficult for the test-taker, the computer adjusts the level in accordance to the test-taker's response. If the level is too easy, the computer adapts the level of difficulty so that the test-taker begins at a higher level (Davies et al., 2005).

New and diverse research methodology has contributed to a holistic perception of working memory. It is thus useful to integrate these methodologies in order to direct future research (Cowan, 1997). Neuroimaging studies have recently been used extensively in working memory and executive processes research and have added much value to the investigation of working memory processes and components (Carpenter et al., 2000). One of the challenges for the interdisciplinary field of working memory research is to be able to combine different kinds of information so that they inform one another. In other words, it is imperative that the theory of working memory informs physiological and neuroimaging investigations into working memory and vice versa (Lockhart, 2000). A recommendation from this study is that future research should attempt to examine working memory from an interdisciplinary position, so as to begin to understand the construct of working memory holistically.

Future research should continue to examine working memory performance in highly educated individuals, in order to begin to understand how working memory deficits present in these subjects. A final recommendation for future research is to begin to address this issue in South Africa, where the rate of traumatic brain injury is high, and where the incidence of working memory deficits in Alzheimer's disease, Parkinson's disease, Schizophrenia, Multiple Sclerosis, HIV and other disorders makes working memory an important and relevant concept in the South African context.

Conclusion

This study examined three core research questions. The first question asked what some of the key issues in computer-based test administration in the South African context were. The study found that computer mouse ability influenced performance on the timed computerised tests, and that computer ability was also linked to prior experience and confidence using a computer, gender and home language. No significant main effects of computer experience, confidence, gender or home language were found when computer mouse ability was removed from the analysis as a covariate. The results of this study supported previous research that suggested that computer familiarity, informed by the social, cultural and historical context of the test-taker, had an effect on performance on computer-based tests (Davies et al., 2005). This suggested that the demographic differences in performance found on the tests may have been informed by experience using a computer rather than reflect true differences in performance between the groups. The results suggested that further exploration of these variables is necessary in the future development and use of computer-based tests, particularly in the South African context.

There were some significant advantages to computerised administration of the tests. First, the study found that the confounding effect of computer familiarity could be largely removed by partialing performance on the computer mouse test. Second, computerised administration improved the standardisation, presentation of material, and scoring of the tests, as well as reducing data handling error and increasing the security of test results. However, the results suggested that computer-based administration may have undermined the validity of some of the tests, particularly on the Digit Span and Stroop Tests. Some of the cognitive demands elicited by the traditional versions of the tests may have been eliminated when the test was administered via computer. Provided that future research establishes the reliability of the tests, that potential threats to validity are clarified, and that computer familiarity is accounted for, computerised testing may play a constructive role in neuropsychological testing in South Africa.

The second and third research questions investigated the cognitive constructs measured by each of the tests administered in this research, and whether the *n*-Back Test demonstrated adequate construct validity as a test of working memory. The study found that the 2-Back condition of the *n*-Back Test correlated significantly with tests that tapped into domain-free executive attention. It correlated significantly with the backward condition of the Digit Span Test, the forward condition of the Spatial Span Test, and part B of the Trail Making Test. However, the 2-Back condition did not load onto the same factor as these subtests, but loaded significantly onto the same factor as the 1- and 3- Back conditions of the *n*-Back Test. The forward and backward conditions of the Digit Span Test loaded significantly onto one factor, and the Stroop and Trail Making subtests loaded significantly onto a separate factor. The forward condition of the Spatial Span Test did not load significantly onto any factor but cross-loaded equally onto the same factor as the *n*-Back, and as the Digit Span Tests.

The results suggested that the tests in this study may have placed demands on different central executive processes, and that the *n*-Back and Digit Span Tests may have loaded onto separate factors according to the type of executive function demanded by each test. Re-examination of the results following a discussion of the effect of computer-based administration of the tests suggested that the *n*-Back Test converged with tests that placed demands on the ability to utilise complex executive attention and diverged from tests that did not place significant demands on executive attention. The results did not provide sufficient evidence to confirm unequivocally that the *n*-Back Test was an adequate measure of working memory ability. However, they did suggest that the *n*-Back task tapped into the central executive component of working memory.

The results presented in this study may not offer up any definitive answers. However, these results should encourage and stimulate enquiry around these issues in a dynamic and evolving South African context.

REFERENCES

Anastasi, A. & Urbina, S. (1997). *Psychological Testing*. New Jersey: Prentice Hall, Inc.

Aron, A., & Aron, E.N. (1999). *Statistics for Psychology (2nd ed.)*. New Jersey: Prentice Hall, Inc.

Baddeley, A. (1982). *Your Memory*. Johannesburg: Flower Press.

Baddeley, A., Bressi, S., Della Salla, S., Logie, R., & Spinnler, H. (1991). The decline of working memory in Alzheimer's Disease: A longitudinal study. *Brain, 114,* 2521-2542

Baddeley, A. (1996a). Exploring the central executive. In G.J. Hitch & R.H. Logie (Eds.), *Working Memory*. Hove: Psychology Press.

Baddeley, A. (1996b, February). *The fractionation of working memory*. Paper presented at the National Academy of Sciences, California, United States of America.

Baddeley, A. (1996c). The Concept of Working Memory. In S. E.Gathercole (Ed.), *Models of Short-term Memory* (pp.1-22). Hove: Psychology Press.

Baddeley, A. (1998a). *Human Memory: Theory and Practice (2nd ed.)*. Boston: Allyn and Bacon.

Baddeley, A. (1998b). Recent developments in working memory. *Current Opinion in Neurobiology, 8,* 234–238

Baddeley, A. (2000a). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A. (2000b). Short-term and working memory. In E. Tulving and F.I.M. Craik (Eds.). *The Oxford Handbook of Memory* (pp. 77-92). New York: Oxford University Press.

Baddeley, A. (2002). Is working memory still working? *European Psychologist, 7,* 85-97

Baddeley, A. (2003). Working memory: looking back and looking forward [Electronic version]. *Nature Reviews/Neuroscience, 4,* 829 – 839.

Banich, M.T. (2004). *Cognitive Neuroscience and Neuropsychology (2nd ed.*). Boston, New York: Houghton Mifflin Company.

Barch, D.M., Braver, T. S., Nystrom, L.E., Forman, S.D., Noll, D.C. & Cohen, J.D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex [Electronic version]. *Neuropsychologia, 10,* 1373-1380.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33,* 570 – 585.

Bohnen, N., Jolles, J., & Twijnstra, A. (1992). Modification of the Stroop Color Word Test improves differentiation between patients with mild head injury and matched controls. *The Clinical Neuropsychologist, 6,* 536-547.

Bower, G.H. (2000). A brief history of memory research. In E. Tulving and F.I.M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 3 -32). New York: Oxford University Press.

Broadbent, D.E. (1958). *Perception and Communication*. Oxford: Pergamon Press.

Broshek, D.K., & Barth, J.T. (2000). The Halstead-Reitan neuropsychological test battery. In G. Groth-Marnat (Ed.), *Neuropsychological Assessment in Clinical Practice (3rd ed.)* (pp. 223-257). New York: John Wiley & Sons, Inc.

Bush, S., Johnson-Greene, D., & Naugle, R. (2002). Interface of information technology and neuropsychology: Ethical issues and recommendations. *The Clinical Neuropsychologist, 16,* 536-547

Carpenter, P.A., Just, M.A., and Reichle, E.D. (2000). Working memory and executive function: evidence from neuroimaging. *Current Opinion in Neurobiology, 10,* 195-199

Callicott, J.H., Mattay, V.S., Bertolino, A., Finn, K., Coppola, R., Frank, J.A., Goldberg, T.E., & Weinberger, D.R. (1999). Physiological Characteristics of Capacity Constraints in Working Memory as Revealed by Functional MRI. *Cerebral Cortex, 9,* 20-26

Chang, L., Speck, O., Miller, E.N., Braun, J., Jovicich, J., Koch, C., Itti, L., & Ernst, T. (2001). Neural correlates of attention and working memory deficits in HIV patients [Electronic Version]. *Neurology, 57,* 1001-1007

Conklin, H.M., Curtis, C.E., Katsanis, J., & Iacono, W.G. (2000). Verbal working memory impairment in schizophrenia patients and their first-degree relatives: Evidence from the digit span task [Electronic Version]. *American Journal of Psychiatry, 157,* 275-277

Conway, A.R.A., Kane, M.J., & Engle, R.W. (2003). Working memory capacity and its relation to general intelligence [electronic version]. *TRENDS in Cognitive Sciences, 7,* 547-552

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user’s guide [Electronic version]. *Psychonomic Bulletin and Review, 12,* 769-786.

Costello, A.B., & Osborne, J.W. (2005). Best practices in Exploratory Factor Analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10,* 1-9. Retrieved 31/01/2008 from http://www.pareonline.net/pdf/vlon7.pdf

Cowan, N. (1997). *Attention and Memory: An Integrated Framework.* New York, Oxford: Oxford University Press.

Davies, C., Foxcroft, C., Griessel, L. & Tredoux, N. (2005). Computer-based and internet-delivered assessment. In Foxcroft, C. & Roodt, G. (Eds.), *An Introduction to Psychological Assessment in the South African Context (2nd ed.)* (pp.176-241). Cape Town, Johannesburg: Oxford University Press.

De Lillo, C. (2004). Imposing structure on a Corsi-type task: Evidence for hierarchical organisation based on spatial proximity in serial-spatial memory. *Brain and Cognition, 55,* 415-426

Dudai, Y. (2002). *Memory from A to Z: Keywords, Concepts and Beyond.* London: Oxford University Press.

Eysenck, M.W. & Keane, M.T. (1995). *Cognitive Psychology: A Student's Handbook.* East Sussex: Psychology Press Ltd.

Fisher, K.N. (2006, February). *Computerised neuropsychological testing in South Africa.* Paper presented at the Tenth Biannual South African Neuropsychological Association Conference, Cape Town, South Africa.

Foxcroft, C.D. (1997). Psychological testing in South Africa: Perspectives regarding ethical and fair practices. *European Journal of Psychological Assessment, 13,* 229-235.

Foxcroft, C. & Roodt, G., (2005). *An introduction to Psychological Assessment in the South African context (2nd ed.).* Johannesburg, Cape Town: Oxford University Press.

Gabrieli, J.D.E., Singh, J., Stebbins, G.T., & Goetz, C.G. (1996). Reduced working memory span in Parkinson's Disease: Evidence for the role of a frontostriatal system in working and strategic memory. *Neuropsychology, 10,* 322-332.

Gazzaniga, M.S., Ivry, B., & Mangun, G.R. (2002). *Cognitive Neuroscience (2nd ed.).* New York: WW Norton and Company

Gibertini, M., & Retzlaff, P.D. (1994). Neuropsychometric issues and problems. In R. Vanderploeg (Ed.), *Clinician's Guide to Neuropsychological Assessment (*pp. 185-209). New Jersey: Lawrence Erlbaum Associates.

Goldberg, E., (2001). *The Executive Brain: Frontal lobes and the Civilised Mind.* New York: Oxford University Press.

Goldman-Rakic, P.S. (1993). Working memory and the mind. *Mind and Brain: Readings from Scientific American Magazine.* New York: W.H Freeman and Company.

Goldstein, G. (1998). Introduction to neuropsychological assessment. In S.R. Beers, G. Goldstein & P.D. Nussbaum (Eds.), *Neuropsychology* (pp.1-5). London, New York: Plenum Press.

Gow, C.A & Stuss, D.T. (1992). "Frontal dysfunction" after traumatic brain injury. *Journal of Neuropsychiatry, Neuropsychology and Behavioural Neurology, 5,* 272-282.

Greene, R.L. (1992). *Human Memory: Paradigms and Paradoxes*. New Jersey: Lawrence Erlbaum Associates.

Grieve, K. W. (2005). Factors affecting assessment results. In In Foxcroft, C. & Roodt, G. (Eds.), *An Introduction to Psychological Assessment in the South African Context (2nd ed.)* (pp.224-241).

Groth-Marnat, G. (2000). *Neuropsychological Assessment in Clinical Practice (3rd ed.).* New York: John Wiley & Sons, Inc

Groth-Marnat, G., Gallagher, R.E., Hale, J.B., & Kaplan, E. (2000). The Wechsler Intelligence Scales. In G. Groth-Marnat (Ed.), *Neuropsychological Assessment in Clinical Practice (3rd ed.)* (pp. 129 - 188). New York: John Wiley & Sons, Inc.

Gur, R.C., Ragland, J.D., Moberg, P.J., Turner, T.H., Bilker, W.B., Kohler, C., Siegel, S.J., & Gur, R.E. (2001). Computerised neurocognitive scanning I: Methodology and validation in healthy people [Electronic version]. *Neuropsychopharmacology, 25,* 766-776.

Hancock, A.B., LaPointe, L.L., Stierwalt, J.A.G., Bourgeios, M.S., & Zwaan, R.A. (2007). Computerized measures of verbal working memory in healthy elderly participants [Electronic version]. *Contemporary issues in communication science and disorders, 34,* 73-85

Harris, P. (1986). *Designing and Reporting Experiments in Psychology*. Philadelphia: Open University Press.

Hausler, J., Sommer, M., & Chroust, S. (2007). Optimizing technical precision of measurement in computerized psychological assessment on Windows platforms [Electronic version]. *Psychology Science, 49,* 116-131

Hazy, T.E., Frank, M.J., & O'Reilly, R.C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience, 139,* 105-118

Health Professions Council of South Africa (2006). South African Guidelines on Computerised Testing. Retrieved 15/04/2008 from http://www.sacna.co.za/html/hspca.html

Howieson, D.B., Loring, D.W., & Hannay, H.J. (2004). Neurobehavioural variables and diagnostic issues. In M.D. Lezak, D.B. Howieson & D. W. Loring (Eds.), *Neuropsychological Assessment (4th ed.).* Oxford, New York: Oxford University Press.

Hulme, C. & Mackenzie, S. (1992). *Working Memory and Severe Learning Difficulties.* Hillsdale: Lawrence Erlbaum Associates.

International Testing Commission. (2005). Guidelines on Computer-based and Internet-delivered Testing. Retrieved 31/01/2008 from http://www.intestcom.org/Downloads/ITC%20Guidelines%20on%20Computer%20-%20version%202005%20approved.pdf

Kane, M.J., Conway, A.R.A., Miura, T.K., & Colflesh, G.J.H. (2007). Working memory, attention control, and the n-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33,* 615-622

Kanji, G, K. (1993). *100 Statistical Tests*. London: Sage Publications.

Kline, P. (1996). *An Easy Guide to Factor Analysis*. London and New York: Routledge

Lezak, M.D. (1995). *Neuropsychological Assessment (3rd ed.).* Oxford, New York: Oxford University Press.

Lezak, M.D (2004). Orientation and Attention. In M.D. Lezak, D.B. Howieson, D.W. Loring (Eds.), *Neuropsychological Assessment (4th ed.).* Oxford, New York: Oxford University Press.

Lockhart, R.S. (2000). Methods of memory research. In E. Tulving and F.I.M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 45-58). New York: Oxford University Press.

Logie, R.H. (1995). *Visuospatial Working Memory.* Sussex: Lawrence Erlbaum Associates.

McAllister, T. (2006a, February). *Neuropsychiatric disorders and mild brain injury.* Paper presented at the Tenth Biannual South African Clinical Neuropsychological Association Conference, Cape Town, South Africa.

McAllister, T. (2006b, February). *Why working memory doesn't work after Traumatic Brain Injury.* Paper presented at the Tenth Biannual South African Clinical Neuropsychological Association Conference, Cape Town, South Africa.

McEvoy, L.K., Smith, M.E. & Gevins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: Effects of memory load and task practice [Electronic version]. *Cerebral Cortex, 8,* 563-574.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative Data Analysis (2nd ed.).* California: SAGE publications, Inc.

Mirsky, A.F., Anthony, B.J., Duncan, C.C., Ahearn, M.B. & Kellam, S.G. (1991). Analysis of the elements of attention: A neuropsychological approach. *Neuropsychology Review, 2,* 109-145.

Nadolne, M.J & Stringer, A.Y. (2000). Neuropsychological assessment: Contexts for contemporary clinical practice. In G. Groth-Marnat (Ed.), *Neuropsychological Assessment in Clinical Practice: A Guide to Test Interpretation and Integration (3rd ed.)* (pp. 26 - 43). New York: John Wiley & Sons, Inc.

Neath, I., & Surprenant, A.M. (2003). *Human Memory (2nd ed.).* California: Wadsworth/Thomson Learning, Inc.

Norman, D.A. (1976). *Memory and Attention: An Introduction to Human Information Processing (2nd ed.).* New York: John Wiley & Sons, Inc.

Nyberg, L. & Cabeza, R. (2000). Brain imaging of memory. In E. Tulving & F.I.M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 501-519). New York: Oxford University Press.

Owen, A.M, McMillan, K.M., Laird, A.R., &Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies [Electronic version]. *Human Brain Mapping, 1,* 46-59.

Paul, R.H., Williams, L.M., & Richard, C.C. (2005). Preliminary validity of "INTEGNEURO™": A new computerized battery of neurocognitive tests [Electronic version]. *International Journal of Neuroscience, 115,* 1549-1567

Pelosi, L., Geesken, J.M., Holly, M., Hayward, M., & Blumhardt, L.D. (1997). Working memory impairment in early multiple sclerosis: Evidence from an event-related potential study of patients with clinically isolated myelopathy. *Brain, 120,* 2039-2058.

Perlstein, W.M., Carter, C.S., Noll, D.C., & Cohen, J.D. (2001). Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *American Journal of Psychiatry, 7,* 1105 - 1113

Phillips, W.A. & Baddeley, A.D. (1989). Learning and memory. In A.D. Baddeley and N.O. Bernsen (Eds.), *Cognitive Psychology: Research Directions in Cognitive Science: European Perspectives Volume One* (pp. 61-83). Hove: Lawrence Erlbaum Associates.

Ponsford, J.L (2000). Attention. In G. Groth-Marnat (Ed.), *Neuropsychological Assessment in Clinical Practice: A Guide to Test Interpretation and Integration (3rd ed)* (pp. 355-392). New York: John Wiley & Sons, Inc.

Postle, B.R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. *Memory and Cognition, 33,* 203-212

Pulos, S. (1997). Construct validity and measures of processing capacity: A response [Electronic version]. *International Journal of Behavioural Development, 20,* 743-745.

Quinette, P., Guillery, B., Desgranges, B., de la Sayette, V., Viader, F., & Eustache, F. (2003). Working memory and executive functions in transient global amnesia [Electronic version]. *Brain, 126,* 1917-1934.

Repvos, G. & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139,* 5-21.

Reynolds, A.G. & Flagg, P.W. (1983). *Cognitive Psychology (2nd ed.).* Boston: Little, Brown & Co.

Rudkin, S., Pearson, D.G., & Logie, R.H. (2007). Executive processes in visual and spatial working memory tasks. *The Quarterly Journal of Experimental Psychology, 60,* 79-100.

Sbordone, R.J. (2000). The executive functions of the brain. In G. Groth-Marnat (Ed.), *Neuropsychological Assessment in Clinical Practice* (pp. 437-453). New York: John Wiley & Sons

Shiffrin, R. & Atkinson, R. (1968). Human memory: A proposed system and its control processes. In K.W. Spence and J.T. Spence (Eds.), *The Psychology of Learning and Motivation (Vol. 2.)* (pp. 115-116). New York: Academic Press.

Sternberg, R.J. (1996). *Cognitive Psychology (2nd ed.)* Fort Worth: Harcourt Brace College Publishers.

Styles, E. A. (1997). *The Psychology of Attention*. East Sussex: Psychology Press Ltd.

Tulving, E. (2000). Concepts of memory. In E. Tulving and F.I.M. Craik (Eds.) *The Oxford Handbook of Memory* (pp. 33-44). New York: Oxford University Press.

Wager, T.D., & Smith, E.E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Journal of Cognitive, Affective and Behavioural Neuroscience, 3,* 255 - 274

Wilde, N.J., & Strauss, E. (2002). Functional equivalence of WAIS-III/WMS-III digit and spatial span under forward and backward recall conditions. *Journal of Clinical Neuropsychology, 16,* 322-330.

Wilde, N.J, Strauss, E., & Tulsky, D.S. (2004). Memory span on the Wechsler scales. *Journal of Clinical and Experimental Neuropsychology, 26,* 539 – 549.

Zillmer, E.A. & Spiers, M.V. (2001). *Principles of Neuropsychology*. United Kingdom: Wadsworth.

Appendix A

Computer-based Questionnaire (Hard Copy)

[Question 1]
Q=How confident are you with using a computer?
1=Confident
2=Not very confident

[Question 2]
Q=Have you had a lot of experience using a computer?
1=Yes
2=No

[Question 3]
Q=How comfortable are you with speaking, writing and reading in English?
1=I am *very* comfortable speaking and reading in English
2=I am *moderately* comfortable speaking and reading in English
3=I am *not very* comfortable speaking and reading in English

[Question 4]
Q=Have you ever been involved in an accident where your head was damaged, you lost consciousness, and you were hospitalised? For example, a motor vehicle accident/sports accident/head trauma
1=Yes
2=No

[Question 5]
Q=Have you ever been treated for any neurological or psychiatric disorder? For example, epilepsy/depression
1=Yes
2=No

[Question 6]
Q=Have you ever been diagnosed as having any kind of learning disability?
1=Yes
2=No

[Question 7]
Q= Did you get enough sleep last night?
1=Yes
2=No

[Question 8]
Q=Do you feel more tired than usual right now?
1=Yes
2=No

[Question 9]
Q=Do you have a headache right now?
1=Yes
2=No

[Question 10]
Q= What is your mood like at the moment?
1=Swings more than usual
2=Stable and same as usual
3=Always low
4=Continually happy

[Question 11]
Q=Do you feel worried and anxious?
1=Yes
2=No

[Question 12]
Q=Are you currently experiencing uncomfortable pain?
1=Yes
2=No

[Question 13]
Q=Are you currently experiencing difficulty with your reading vision?
1=Yes, I AM NOT wearing my glasses/contact lenses
2=No, I AM wearing my glasses/contact lenses
3=No, I do not wear glasses/contact lenses

Appendix B

Presentation of materials via computer

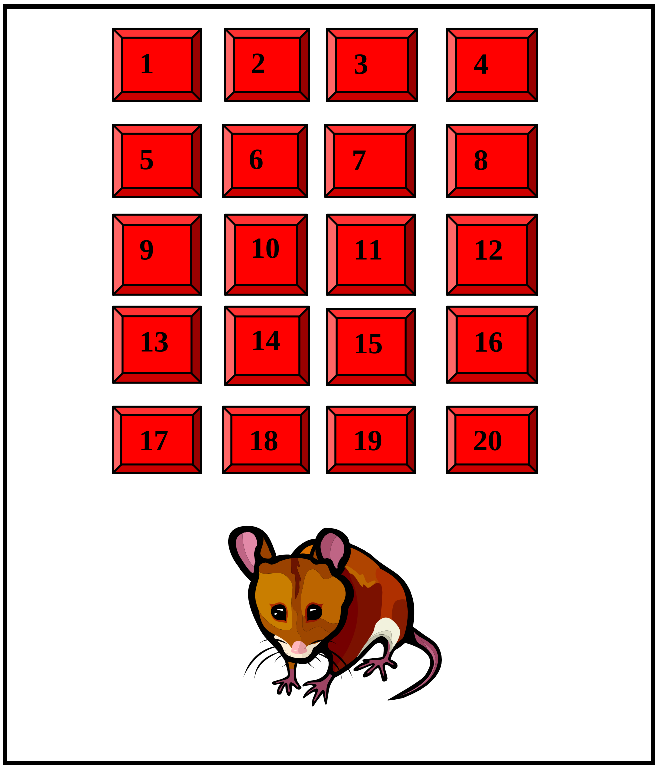


Figure 1. Visual example of part A of the Computer Mouse Test item presentation

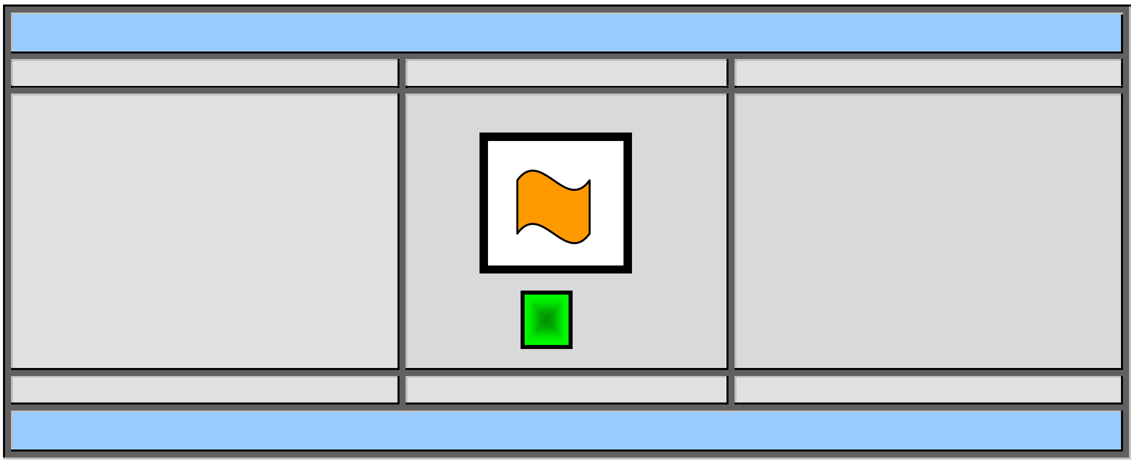


Figure 2. Visual example of the N-back Test item presentation

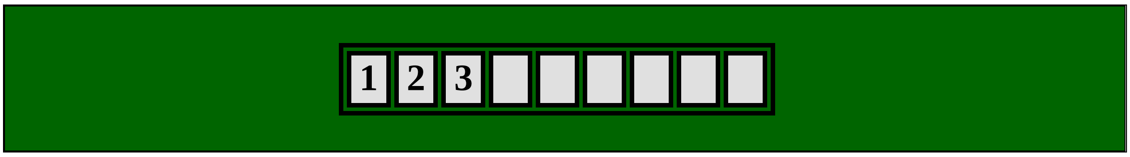


Figure 3. Visual example of the Digit Span Test response screen

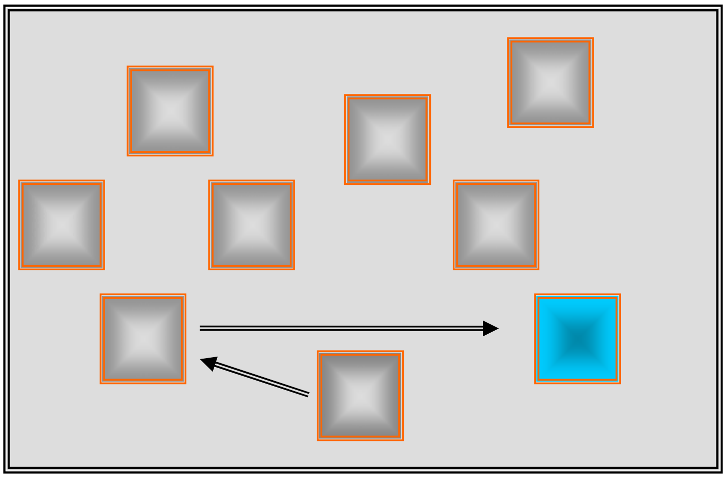


Figure 4. Visual example of the Spatial Span Test item presentation (note: arrows denote movement of the sequence)

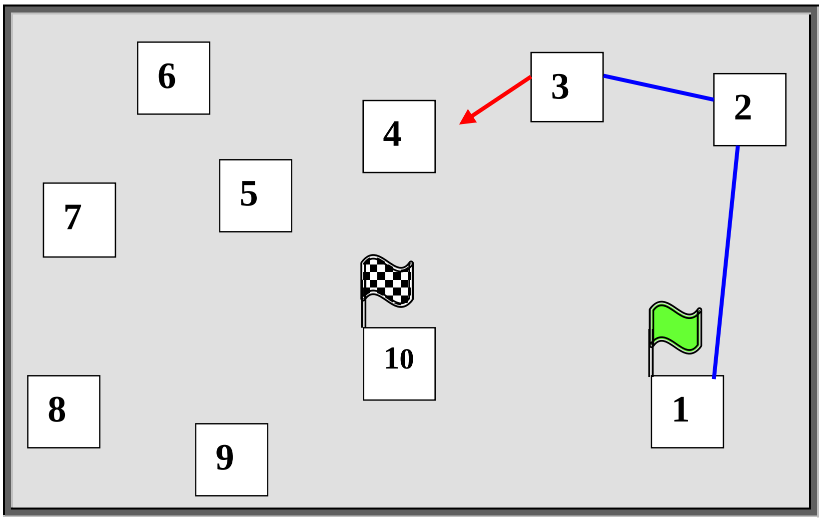


Figure 5. Visual example of the Trail Making Test item presentation

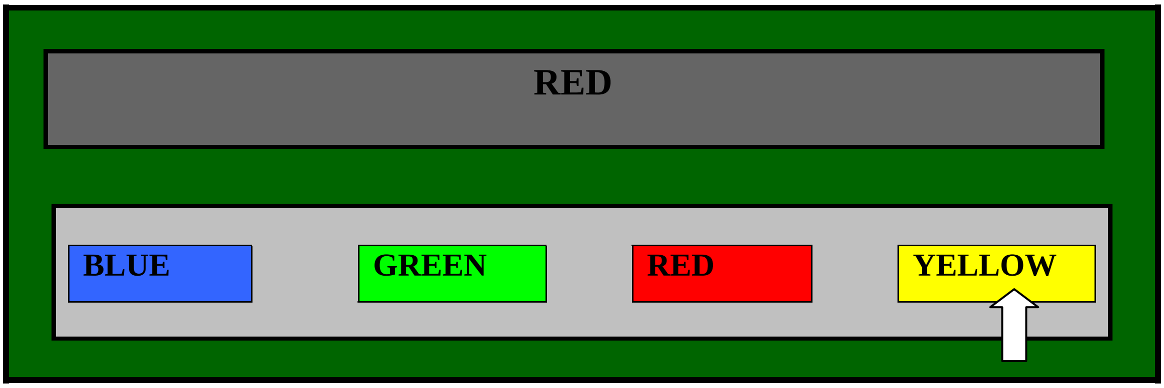


Figure 6. Visual example of the Stroop Test (interference phase) item presentation

Appendix C

Information Sheet

My name is Katharine Fisher and I am conducting research for the purpose of obtaining a Masters degree in Research Psychology. My area of interest is memory. Working memory describes the process by which we are able to hold information on-line in order to update this information for cognitive and behavioural guidance. It allows for continuity of experience and is therefore integral to complex and adaptive human functioning. Neuropsychologists in South Africa do not have enough information about how working memory operates in educated South African adults. Part of this research will explore how working memory is broken down into different levels of processing. In addition to this, this research aims to explore how variables such as age, gender, and English proficiency affect performance on currently utilised computer-based neuropsychological tests.

I would like to invite you to participate in this study. Participation in this research will entail completing a brief computer-based questionnaire, a computer mouse test, and four computer-based neuropsychological tests of working memory. The testing process will take up to an hour to complete. Participation is voluntary, and no person will be advantaged or disadvantaged in any way for choosing to complete or not to complete the tests. The interview itself will not be recorded, but computer-based results will be documented. Your completed tests will not be seen by anyone but myself, and a qualified neuropsychologist assisting in the interpretation of test results. Your test results will only be examined in relation to all other test results. This means that feedback given to you will be in the form of group results and no individual test results will be released. Your responses will be kept confidential, and no information that could identify you will be included in the report. Raw data (including identifying data) will be destroyed once the research is marked and the results have been returned to me. You may refuse to answer any questions that you would prefer not to answer, and you may choose to withdraw from the study at any point.

If you choose to participate in this study please fill in your name in a time slot that is convenient for you. You may contact me to ask any questions about the research, or to move or cancel an appointment. Your participation in this study is greatly appreciated. This research will contribute to both a larger body of knowledge on computer-based neuropsychological testing, as well as to the University [redacted] and to neuropsychologists currently practising in the South African context.

Sincerely,

Katharine Fisher

Cell phone: [redacted]
Email: [redacted]

Appendix D

Table 10.

Tests for a Normal Distribution

Tests For Normality

Variable	Skewness	Kurtosis	Kolmogorov-Smirnov *D*
Stroop1	0.62	0.71	0.08
Stroop2	0.23	-0.20	0.05
Stroop3	0.59	-0.12	0.09
Stroop4	0.25	-0.49	0.09
0Back	-	-	-
1Back	-0.93	0.48	0.18 *
2Back	-0.31	-0.20	0.10 *
3Back	0.31	-1.62	0.32 *
Digits F	0.06	-0.54	0.17 *
Digits B	0.03	0.28	0.16 *
Spatial F	0.29	-0.49	0.16 *
Trails A	0.82	0.75	0.09
Trails B	-0.13	-0.70	0.06
Mouse Test	1.03	1.18	0.11 *

*p < .01

Appendix E

Significant differences in tests scores between the groups (p < .015)

Table 11.
Mean scores males and females on the tests

Subtest	Gender Male *Mean/Standard Deviation*	Gender Female *Mean/Standard Deviation*
Stroop1	*M* = 82.78 (*SD* = 11.14)	*M* = 89.06 (*SD* = 9.33)
Stroop2	*M* = 75.65 (*SD*= 9.18)	*M* = 81.64 (*SD* = 8.52)
Mouse Test	*M* = 33.08 (*SD* = 6.42)	*M* = 37.83 (*SD* = 7.39).

Table 12.
Mean scores of subjects with 12-15 years of education and subjects with 16-18 years of education on the tests

Subtest	12-15 Years Education *Mean/Standard Deviation*	16-19 Years Education *Mean/Standard Deviation*
1 Back	*M* =18.27 (*SD* = 1.72).	*M* = 17.09 (*SD* =1.92).

Table 13.

Mean scores of first and second language English speakers on the tests

Subtest	1st Language English *Mean/Standard Deviation*	2nd Language English *Mean/Standard Deviation*
Stroop1	*M* = 85.79 (*SD* = 9.58)	*M* = 91.41 (*SD* = 10.4)
Stroop2	*M* = 78.63 (*SD* = 8.87)	*M* = 83.69 (*SD* = 8.51)
Digits B	*M* = 5.82 (*SD* = 1.13)	*M* = 5.15 (*SD* = 1.32)
Trails A	*M* = 31.44 (*SD* = 6.90)	*M* = 36.88 (*SD* = 9.61)
Trails B	*M* = 49.20 (*SD* = 11.3)	*M* = 56.08 (*SD* = 10.1)

Table 14

Mean scores of subjects who were confident using a computer and not confident using a computer on the tests

Subtest	Confident *Mean/Standard Deviation*	Not Confident *Mean/Standard Deviation*
Stroop3	*M* = 86.55 (*SD* = 9.76).	*M* = 91.3 (*SD* = 10.91)

Table 15

Mean scores of subjects who were experienced using a computer and not experienced using a computer on the tests

Subtest	Experienced *Mean/Standard Deviation*	Not Experienced *Mean/Standard Deviation*
Stroop1	*M* = 85.49 (*SD* = 8.79)	*M* = 94.69 (*SD* = 11.46).
Stroop2	*M* = 78.55 (*SD* = 8.29)	*M* = 86.09 (*SD* = 9.37).
Stroop3	*M* = 89.19 (*SD* = 11.80)	*M* = 98.58 (*SD* = 15.05)
Trails A	*M* = 31.94 (*SD* = 7.26).	*M* = 37.49 (*SD* = 10.0)
Mouse Test	*M* = 35.17 (*SD* = 6.53	*M* = 41.86 (*SD* = 8.15)

Appendix F

Table 16.

Correlation Matrix

Subtest	**Stroop1**	**Stroop2**	**Stroop3**	**Stroop4**	**1Back**	**2Back**	**3Back**	**Digits F**	**Digits B**	**Spatial F**	**Trails A**	**Trails B**	**Mouse**
Stroop1	1.00	.84**	.79**	.66**	.03	-.15	-.04	-.11	-.19	-.06	.57**	.40**	.58**
Stroop2	.84**	1.00	.84**	.70**	-.02	-.18	-.10	-.07	-.11	-.15	.51**	.38**	.54**
Stroop3	.79**	.84**	1.00	.75**	-.09	-.16	-.00	-.04	-.13	-.10	.47**	.35**	.46**
Stroop4	.66**	.70**	.75**	1.00	-.04	-.28**	-.06	-.10	-.21*	-.13	.43**	.42**	.37**
1Back	.03	-.02	-.09	-.04	1.00	.54**	.23*	-.02	.12	.12	.00	-.11	.00
2Back	-.15	-.18	-.16	-.28**	.54**	1.00	.29**	.15	.32**	.258*	-.24*	-.33**	-.16
3Back	-.04	-.10	-.00	-.06	.23*	.29**	1.00	-.04	.05	.17	-.09	-.14	-.10
DigitsF	-.11	-.07	-.04	-.10	-.02	.15	-.04	1.00	.43**	.22	-.196*	-.12	-.06
DigitsB	-.192	-.11	-.13	-.21*	.12	.32**	.05	.43**	1.00	.12	-.23*	-.11	-.12
SpatialF	-.06	-.15	-.10	-.13	.12	.258*	.17	.22*	.12	1.00	-.11	-.14	-.08
TrailsA	.57**	.51**	.47**	.43**	.00	-.24*	-.09	-.196*	-.23*	-.11	1.00	.59**	.43**
TrailsB	.40**	.38**	.35**	.42**	-.11	-.33**	-.14	-.12	-.11	-.14	.59**	1.00	.24*
Mouse	.58**	.54**	.46**	.37**	.00	-.16	-.10	-.06	-.12	-.08	.43**	.24*	1.00

Note: * p < .015; ** p < .0001

Appendix G

```
4.0 +
   |
   |             1
   |
3.5 +
   |
   |
   |
3.0 +
   |
   |
   |
2.5 +
   |
   |
   |
2.0 +                    2
   |
   |
   |
1.5 +
   |                             3
   |
   |                                   4
1.0 +
   |                                            5
   |                                                  6
   |
0.5 +                                                       7       8
   |                                                                      9
   |                                                                            10
   |                                                                                   11
   |                                                                                         12
0.0 +
   --------+------+------ +------+------+------+------+------+------+------+------+------+------+--
           0      1      2       3      4      5      6      7      8     9      10     11     12
```

Figure 7. Cattell’s Scree Plot of the Eigenvalues

Appendix H

Table 17.

Residual Correlation Matrix

Subtest	Stroop1	Stroop2	Stroop3	Stroop4	1Back	2Back	3Back	DigitsF	DigitsB	SpatialF	TrailsA	TrailsB
Stroop1	0.22	-0.00	-0.04	-0.10	-0.03	-0.00	-0.05	-0.01	-0.04	0.02	-0.05	-0.08
Stroop2	-0.00	0.20	-0.00	-0.07	-0.00	0.00	-0.04	-0.06	-0.01	-0.08	-0.10	-0.12
Stroop3	-0.04	-0.00	0.20	0.00	-0.08	0.00	0.05	-0.04	-0.06	-0.02	-0.12	-0.15
Stroop4	-0.10	-0.07	0.00	0.33	0.02	-0.02	0.01	-0.00	-0.04	0.00	-0.14	-0.07
1Back	-0.03	-0.00	-0.08	0.02	0.37	-0.05	-0.26	0.07	0.04	-0.13	0.05	0.10
2Back	-0.00	0.00	0.00	-0.02	-0.05	0.28	-0.18	-0.02	-0.01	-0.12	0.03	0.05
3Back	-0.05	-0.04	0.05	0.01	-0.26	-0.18	0.59	0.08	0.03	-0.01	-0.04	0.03
DigitsF	-0.01	-0.06	-0.04	-0.00	0.07	-0.02	0.08	0.28	-0.19	-0.03	0.12	0.07
DigitsB	-0.04	-0.01	-0.06	-0.04	0.04	-0.01	0.03	-0.19	0.42	-0.17	0.12	0.16
SpatialF	0.02	-0.08	-0.02	0.00	-0.13	-0.12	-0.01	-0.03	-0.17	0.76	0.09	0.07
Trails A	-0.05	-0.10	-0.12	-0.14	0.05	0.03	-0.04	0.12	0.12	0.09	0.56	0.19
Trails B	-0.08	-0.12	-0.15	-0.07	0.10	0.05	0.03	0.07	0.16	0.07	0.19	0.61

APPENDIX I

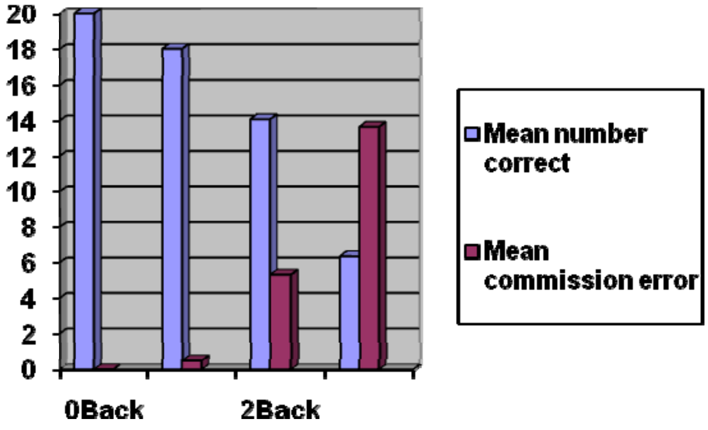


Figure 8. Mean correct scores and mean commission error on the *n*- Back Test

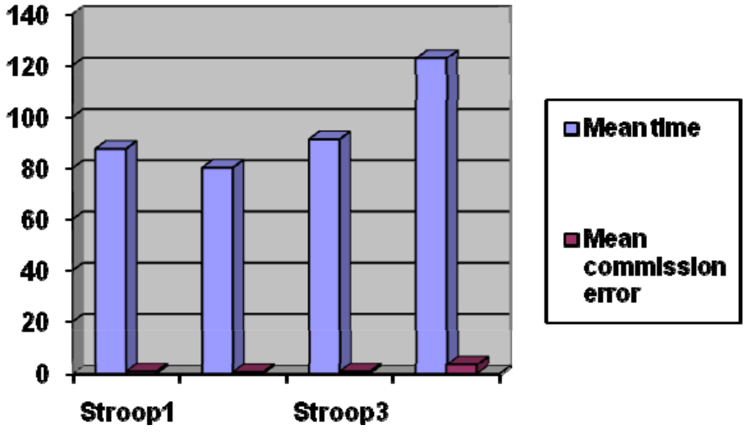


Figure 9. Mean time and mean commission error on the Stroop Test

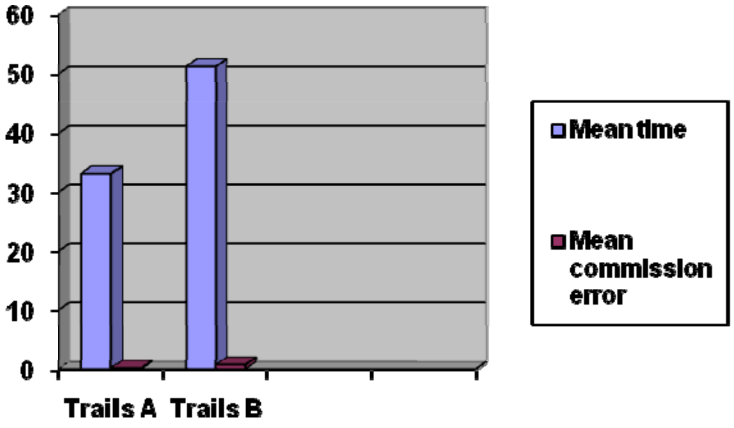


Figure 10. Mean time and mean commission error on the Trail Making Test