Production of Australian normative data for an alternate scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised

Helen Ann Jeges, Ba.App.Sci. (Hons)

Victoria University, School of Social Sciences and Psychology,
Faculty of Arts, Education and Human Development, Melbourne, Australia

Research thesis submitted as partial fulfilment of the requirements for the degree of Doctor of Psychology (Clinical Neuropsychology)

1st May 2012

ABSTRACT

Accurate assessment of memory functioning is integral in the framework of comprehensive neuropsychological examination. Knowledge of the integrity of memory functioning contributes to decisions regarding diagnoses, competency, rehabilitation and surgery. The Wechsler Memory Scale-Revised is a frequent memory assessment tool in everyday neuropsychological practice. The present study reports the modification and validation of an alternate scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised and generation of Australian normative data. This test is frequently used in clinical practice and therefore an improved scoring method and collation of locally appropriate normative data forms a significant contribution to knowledge and extension of the evidence base for the practice of clinical neuropsychology. An alternate scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised was previously developed and preliminarily validated in a clinical population by Dowling and Clark (2000). The current research project modified the Alternate Scoring System to improve reliability. Subsequent to its modification, normative data was required to enhance clinical utility in an Australian population. The scoring system was applied to subject responses from a large-scale normative study conducted between 1996 and 1998. The Macquarie University Neuropsychological Normative Study involved 399 Australian adults aged between 18 and 34 years of age and included administration of the Wechsler Memory Scale-Revised. As hypothesised, the alternate scoring system generated a similar grading of memory to the original scoring system. Major outcomes of the present study include a substantial literature review to support the development of the alternate scoring system, modification of the alternate scoring system in collaboration with original authors, further validation of the alternate scoring system through comparison with the original scoring system, and production of Australian normative data. Implications of the modified alternate scoring system, together with the normative data produced through the present study, have broadened the clinical applications of the WMS-R Visual Reproduction subtest and provided another tool to guide result interpretation in the assessment of visual memory functioning in an Australian population.

DECLARATION

I, Helen Ann Jeges, declare that the Doctor of Psychology (Clinical Neuropsychology) thesis entitled 'Production of Australian normative data for an alternate scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised' is no more than 40,000 words in length including quotes and exclusive of tables, figures, appendices, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Helen Ann Jeges
1st May 2012

ACKNOWLEDGMENTS

Firstly, I would like to thank my principal supervisor Associate Professor Gerard Kennedy, for his ongoing support, assistance, and encouragement; without which, this work would not have been completed.

I would also like to express my gratitude to my initial supervisor, Dr Peter Dowling, who provided me with a source of inspiration and never-ending enthusiasm, and above all else, for constantly reassuring me that I could do it.

I also wish to thank Dr Arthur Shores and Dr Jane Carstairs, from the Department of Psychology, Macquarie University, Sydney, Australia, for providing me with access to the data collected through the Macquarie University Neuropsychological Normative Study (MUNNS).

I owe a great debt of thanks to my colleague, Dr Areti Plitas, Clinical Neuropsychologist, for generously sharing her time, expertise and encouragement.

Thanks also to Associate Professor Gavin Ivey for taking the time to review and comment on the research, prior to submission.

Lastly, I wish to convey the full depth of my appreciation to my husband and family for their continued interest in my research, the unconditional support they offered me, and for always listening to my thoughts.

TABLE OF CONTENTS

ABBREVIATIONS

AD	Alzheimer's Disease
ASS	Alternate Scoring System
BVMT	Brief Visuospatial Memory Test
BVRT	Benton Visual Retention Test
DLPFC	Dorsolateral Prefrontal Cortex
ERP	Evoked Response Potential
FM	Figural Memory
fMRI	Functional Magnetic Resonance Imaging
FSIQ	Full Scale Intelligence Quotient
HD	Huntington's Disease
LM	Logical Memory
LTM	Long-term Memory
LTP	Long-term Potentiation
MEG	Magnetoencephalography
MFD	Memory for Designs Test
MQ	Memory Quotient
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MUNNS	Macquarie University Neuropsychological Normative Study
NMDA	N-methyl-D-aspartate
OSS	Original Scoring System
PET	Positron Emission Tomography
ROCFT	The Rey-Osterrieth Complex Figure Test
STM	Short-term Memory
TBI	Traumatic Brain Injury
VePA	Verbal Paired Associates
ViPA	Visual Paired Associates
VMS	Visual Memory Span
VR	Visual Reproduction
WAIS-R	Wechsler Adult Intelligence Scale-Revised
WMS	Wechsler Memory Scale
WMS-R	Wechsler Memory Scale-Revised

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This introduction provides an overview of the concepts of memory and its assessment in the Australian context. Using a selective historical review of literature, this overview aims to detail the evolution of understanding about the assessment of memory and further, to identify limitations of existing memory assessments. The introduction concludes with the rationale for the production of Australian normative data for an alternate scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised and subsequent hypotheses which this study aimed to answer.

1. Memory

Memory processes form an essential component of cognitive functioning and have been described, in the seminal work of Squire, as "the persistence of learning in a state that can be revealed at a later time" (1987, p.3). Learning broadly involves the acquisition of new information, or knowledge, and is necessary for memory to occur; therefore, learning and memory are thought to represent the constant adaptations of neural connections to the environment (Eichenbaum, 2002).

1.1 Memory systems

The fundamental processes involved in memory function and the structure of memory have long been debated by researchers holding different views. Historically, the concept of distinct memory systems has been around since before the twentieth century (Squire, 2004). The following processes have generally been accepted to occur in the various theories that have been proposed about the nature of memory: (1) attention and orientation; (2) encoding, referring to the moving of information into a more permanent memory store through either acquisition, via a sensory buffer or consolidation into longer-term representation; and (3). retrieval, which involves moving information from memory store to 'awareness' in a conscious sense (Craik & Lockhart, 1972; Tulving & Thompson, 1973). The stages of encoding, storage and retrieval are not mutually exclusive and complex interactions exist between them. For example, the quality of encoding affects both storage and retrieval; with information generally better recalled under conditions that are similar to when it was learned (i.e., context-dependent memory), and repeated retrieval of information can increase the probability of it being retrieved at a later time (Bear, Connors, & Paradiso, 1996; Emilien, Durlach, Antoniadis, Van der Linden, & Maloteeaux, 2004).

1.2 Taxonomy of memory

Currently, the basic concepts of memory are stated to be that of 'declarative' (i.e., explicit) and 'nondeclarative' (i.e., implicit). Conscious, declarative memory can be defined as the capacity to consciously recollect conscious information related to facts, locations, and personal information, in addition to information that has been processed in some manner. This is also known as episodic memory and it allows us to recall our past experiences (Tulving, 2002; Squire, 2009a; Squire, 2009b). Nondeclarative memory includes unconscious information such as skills, procedures, habits and within this memory; information is learned in the same way in which it was encountered (Squire, 1992; Tulving, 1985). There is agreement amongst researchers that the processes of memory are likely to be different for declarative and nondeclarative memory (Squire & Knowlton, 1994). Memory has also been classified into several groups or subtypes. Squire (1987) classifies memory subtypes by modality into the following five subtypes: verbal, including words and stories; visual, comprising of pictures, faces and also music; spatial and kinaesthetic, involving similar processes to visual memory; emotion and feelings, usually connected with other percepts; and multi-modal versus uni-modal memory.

Memory is also often classified into the following temporal subtypes: sensory or registration memory, this lasts milliseconds to seconds and is longer for auditory than visual memory; short-term memory (STM), which lasts between seconds and hours and is related to working memory; and long-term memory (LTM), which can last for days, months and years (Atkinson & Shiffrin, 1968; Tulving, 1972). This distinction is somewhat similar to the distinction made between recent and remote memory (Deweer, Pillon, Pochon, & Dubois, 2001). These two memory subtypes make reference to memory that occurs in relation to time in the present. Some types of memory problems have a 'temporal' gradient, with this being more severe for recent memory (Eichenbaum, 2002; Parkin, 1997; Squire, 2009a).

1.3 Models of memory

1.3.1 Multi-store memory model

In 1968 an influential multi-store model of human memory was proposed by Atkinson and Shiffrin. This model posited two distinct 'memory stores', which were: short-term memory (STM) and long-term memory (LTM). Later, a third memory store was added to this model, which was specifically: 'sensory memory'.

Short-term memory is also often referred to as ‘working memory’, and is the type of memory where information is temporarily stored within neural connections. The duration of STM was initially thought to be limited to eighteen to twenty seconds (Peterson & Peterson, 1959). However, more recently Cohen and Eichenbaum (1993) have argued that STM can endure for a period of up to several hours. The capacity of STM was famously reported by Miller (1956) in a paper where he demonstrated that about seven pieces, or ‘chunks’, (+/- two) of information could reside in STM at the same time. Once the information has entered STM, it fades or decays; a process expedited when new information enters the limited store. Unless rehearsed, the temporary store of information in STM is soon lost as attention is focussed in other directions. STM has been depicted as the ‘bottleneck’ of the human information processing system because the capacity of STM limits the amount of information that can be held in mind simultaneously and also on the duration for which it can be held once attention is withdrawn (Baddeley, 1995). To retain information in STM, it is often encoded verbally, although other strategies may also be used such as visualisation. These strategies make it possible to ‘rehearse’ the information, thereby allowing it to enter long-term memory (Parkin, 1997).

Long-Term Memory is a memory store in the human brain that is considered to be relatively permanent. The properties of LTM vary significantly from that of STM, its capacity is thought to be virtually unlimited, it can endure for a lifetime and humans do not appear to exhaust the capacity for storage of new information even after a full lifetime (Atkinson & Shiffrin, 1968). There is no clear consensus suggesting an accurate method of determining how long memories can be stored in LTM. Permanent losses of information in LTM are known to result from brain damage (Zola, 1997).

LTM has three basic processes; storage, deletion and retrieval. In the storage of LTM it appears that information is organised according to meaning and is associatively linked (Eichenbaum, 2002). There are two types of retrieval of information in the LTM process which are those of: recall and recognition. Recall reproduces the information from memory, while in recognition the information is presented providing the knowledge that the information has been known or seen before. Recognition memory is considered to be less complex than memory recall (Buffalo, Reber, & Squire, 1998).

Sensory memory is actually several memory systems, which are each associated with a specific sense. In other words, there is a sensory memory for vision, which is referred to

as ‘iconic’ memory and another for auditory input referred to as ‘echoic’ memory. (Atkinson & Shiffrin, 1968). These two memory systems share the characteristic of pure raw perceptual processing with no additional processing; but differ in capacity and duration. Iconic memory is purported to last approximately half to one second, whereas echoic memory was proposed to be more enduring lasting for four to five seconds (Baddeley 1994; Sperling, 1960).

1.3.2 Model of working memory

In 1974, Baddeley and Hitch modified the notion of STM by proposing a model of working memory that included the idea that two ‘slave systems’ serve long-term memory: ‘phonological loop’ and ‘visuospatial sketchpad’. The two systems were argued to temporarily store information, as well as perform operations, such as rehearsal, that would maintain information and subsequently allow it to be transferred to LTM. The third component of the working memory model was the ‘central executive’ which provided an interface between the phonological loop, visuospatial sketchpad, and LTM. In addition to providing traditional frontal lobe functions the central executive allocated attention to different processes and carried out different activities, such as organisation. The working memory system has, more recently, been purported to serve a broader role in the temporary maintenance and manipulation of complex cognitive processes such as comprehension, learning and reasoning (Emilien et al., 2004; Lezak, Howieson, & Loring, 2004).

1.3.3 Summary of memory processes

When reviewing evolution of the understanding of memory processes, the literature indicates that essential elements of memory systems include: attention, encoding, storage and retrieval. Conceptually, memory has been classified both in a declarative or nondeclarative dichotomy, and segregated by temporal subtype, such as, STM and LTM. Cognitive models of memory function, including: the multi-store model of human memory (Atkinson & Shiffrin, 1968) and the model of working memory (Baddeley & Hitch, 1974) can be better understood with an exploration of the neuroanatomy and neurophysiological correlates of memory.

1.4 The neuroanatomy of memory

Memory is not a unitary faculty of the mind and no single brain structure accounts for all memory function (Squire, 2004; Squire, 2009b). The physical representation or location of a memory is often referred to as a memory trace, or 'engram'. Engrams are

widely distributed among the connections that link the cells of a memory system and include the same neurons that are involved in sensation and perception. The brain can utilise the same cortical area for both processing of sensory information and storage of memories (Eichenbaum, 2002).

Studies investigating the neural and structural bases of typical memory function in the years before the emergence of modern neuroimaging techniques were limited mainly to the examination of brain damaged patients (Squire, 1986; Squire, 2009b), derived from animal models (Mishkin, 1978; 1982) and from ablation studies in animals (Damasio, Graff Radford, Eslinger, Damasio, & Kassell, 1985; Zola-Morgan, Squire, Amaral, & Suzuki, 1989. The availability of modern neuroimaging technology, such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), Event Related Potentials (ERP) and Magnetoencephalography (MEG), has substantially contributed to knowledge regarding the structural and functional aspects of typical memory function (Mayes & Montaldi, 2001; Squire, Amaral, & Press, 1990). The cumulative wealth of this contemporary research clearly indicates that, although no single brain structure or cellular mechanism accounts for all learning, all cortical areas do not contribute equally to memory (Cabeza & Nyberg, 2000). The overwhelming consensus reached across numerous, methodologically sound studies, concludes that memory processes are mediated by multiple inter-related and connected neural systems (Squire, 2004).

Consistent evidence has been presented that particular parts of the brain are involved in diverse forms of memory and that several different aspects of learning and memory utilise distinct neural systems (Squire & Knowlton, 1994). For example, short-term memories are modality specific and consequently, involve a range of modality specific storage areas in the brain. Various components of brain systems, such as medial temporal lobe structures, are involved in storing particular types of information by means of synaptic mechanisms. Although distinct neural systems subserve varying forms of memory function, interconnectedness remains a prime feature through the connections between cortical areas, as evidenced in a series of fMRI studies (Figueiredo et.al. 2008; Alessio et al, 2011).

1.4.1 Medial temporal lobes

Evidence for multiple memory systems from individual human case studies and animal studies indicated distinct neural systems for impairments of declarative memory

(Mishkin, 1978). Using an animal model of memory impairment in monkeys, Mishkin (1982) was able to describe a detailed, human model of limbic memory function. His series of neuroanatomical studies, combined with behavioural observations assisted in the identification of the key anatomical components of the medial temporal lobe that were proposed to support and subserve declarative memory function (Mishkin, Malamut, Bachevalier, 1984).

The role of the medial temporal lobe in declarative memory formation has been outlined by Cohen and Squire (1980) and although there appears to be a division of labour within the medial temporal lobe, the available data do not support simple dichotomies, such as nondeclarative versus declarative memory. More recently, Cohen and Eichenbaum (1993) proposed a theory of relational memory which argues that as sensory information enters the medial temporal lobe and is processed; it leads to the storage of memories in a manner that relates all of the things happening at the time the memory was stored with the memory itself.

It has also been confirmed that a group of interconnected limbic structures in the medial temporal lobe play a critical role in the consolidation of declarative memory and recognition memory (Squire & Zola-Morgan, 1991; Clark & Squire, 2010). Anatomically related structures of the medial temporal lobe are principally concerned with memory and operate within the neocortex to establish and maintain long-term memory (Squire, Stark, & Clark, 2004). Lesions in this area typically result in amnesia of varying severity.

A famous case study of patient H.M. who underwent a bilateral medial temporal lobe resection which resulted in severe anterograde amnesia (i.e., an impairment in new learning ability) and retrograde amnesia (i.e., impaired recollection for events in the years prior to the surgery), added further evidence to the role of the medial temporal lobe in memory (Scoville & Milner, 1957). H.M.'s amnesia occurred against a background of mostly intact intellectual and perceptual functions, indicating that memory could be a distinct cerebral function that was separable from other cognitive abilities. The case of H.M. was also principally revealing, as H.M. retained relatively intact immediate and remote memory abilities, demonstrating that the structures damaged were not the definitive depository of memory (Squire, 2009a; Squire, 2009b). Animal research involving comparable lesions approximating the loss sustained by

patient H.M. reproduced similar features of memory impairment in monkeys (Zola-Morgan & Squire, 1993).

1.4.2 Hippocampus

Of all medial temporal lobe structures, interest in the critical role of the hippocampus in memory dates from the earliest studies of memory function in brain damaged patients (Adeyemo, 2002; Squire, 2009b). The primary function of the hippocampus is to encode information into STM, which also necessarily involves working memory processes. It is also involved in retrieval of information from long-term memory (de Haan, Mishkin, Baldeweg, & Vargha-Khadem, 2006).

As expected, given the predilection of the medial temporal lobe for the processing and storage of declarative memory; memory deficits following hippocampal lesions are not global, but are much more specific, predominantly involving declarative or declarative memory (Eichenbaum, 2004). The evidence from H.M and animal models strongly implicated the hippocampal complex (including the CA fields, the dentate gyrus, and the subicular complex) as vital structures for the acquisition of new episodic and semantic memories (Mishkin, 1982; Tulving & Markowitsch, 1998; Broadbent, Clark, Zola & Squire, 2002; Deweer et.al, 2001; Lee, Yip, & Jones-Gotman, 2002). Even partial damage to the hippocampus has been found to produce substantial memory impairment in humans and monkeys (Zola-Morgan & Squire, 1993; Mishkin, Vargha-Khadem, & Gadian, 1998; Winocur, Moscovitch, Caruana, & Binns, 2005).

The hippocampal complex in each hemisphere has been postulated to be specialised for different types of declarative memory (Milner, 1971). The left hippocampi processes language and verbally mediated information, while the right hippocampi is reported to process visually based and other sensory information (Desgranges, Baron, & Eustache, 1998; Squire & Butters, 1992). Indeed, lesions of the structures in the left mesial temporal lobe has been shown to consistently impair verbal memory and learning (Chelune, Naugle, Luders, & Awad, 1991; Frisk & Milner, 1990; Herman, Connell, Barr & Wyler, 1995; Ivnik, Sharbrough, & Laws, 1987; Rausch & Babb, 1993; Saling et al., 1993).Whereas, lesions of the structures in the right mesial temporal lobe has been shown to impair the learning, and delayed recall, of visual and spatial information (Helmstaedter, Pohl, Hufnagel & Elger, 1991 ; Milner, 1965; Morris, Abrahams, & Polkey, 1995; Piguet, Saling, O'Shea, Berkovic, & Bladin, 1994; Smith & Milner, 1989; Trenerry, Jack, Cascino, Sharbrough, & Ivnik, 1996).

Results from a study of patient R.B., who had circumscribed memory impairment as a result of a hypoxic brain injury, revealed that a bilateral lesion involving the CA1 field of the hippocampus was sufficient to produce severe anterograde amnesia (Zola-Morgan, Squire, & Amaral, 1986), although concurrent damage to the entorhinal, perirhinal and parahippocampal cortices is reported produce a more severe amnesia (Mishkin, 1978; Squire, 1986; Zola, 1997). Lesions that extend to the entorhinal and perirhinal cortices in association with the hippocampus are implicated in a severe retrograde amnesia extending several decades (Emilien et al, 2004; Rempel-Clower, Zola, Squire, & Amaral, 1996; Baxter & Murray, 2001; Corkin, Amaral, Gonzalez, Johnson, & Hyman, 1997; Zola-Morgan et al, 1986; Markowitsch, 2000).

The fornix is a bundle of output axons that link the hippocampus to the thalamus and hypothalamus. Lesions to the fornix have been shown not to affect procedural learning, but do affect declarative memory formation. Furthermore, lesions to the fornix have not been shown to produce long lasting memory impairment in monkeys (Zola-Morgan et al., 1989).

Neuroimaging studies also consistently indicate a reduction in hippocampal size in those with clinically significant memory impairments (Squire, Amaral, & Press, 1990; Markowitsch, 2003; Mayes & Montaldi, 1999). Quantitative MRI studies have indicated the role of hippocampal atrophy in memory dysfunction in Alzheimer's disease (Deweer et al., 1995) and following traumatic brain injury (Bigler et al., 1996). In addition to the hippocampus, other specific temporolimbic structures implicated in memory function include the parahippocampal regions and the amygdala (Kelley et al., 1998; McGaugh et al., 1993; Tranel & Damasio, 1995; Squire & Zola-Morgan, 1991; Zola-Morgan & Squire, 1993).

1.4.3 Parahippocampal region

The parahippocampal region consists of two distinct and adjacent cortical zones: the parahippocampal cortex and rhinal cortex. These two areas are viewed as highly interconnected cortical zones that, as a whole, receive and accumulate input from multiple neocortical association areas and send cortical information to the hippocampus (Zola-Morgan et al., 1989; Suzuki & Amaral, 1994). The parahippocampal region constitutes the main convergence site for neocortical input to the hippo0uocampus for declarative memory formation. Input connections to the parahippocampal region arise from virtually every higher-order association area, such as pre-frontal, parietal and

temporal. Only highly pre-processed sensory information reaches the medial temporal lobe structures (Kelley et al., 1998). The parahippocampal region of the brain plays a crucial role, in conjunction with other close structures in the brain, in functions such as learning, memory, emotions, and more multifaceted behavioural processes (Squire & Knowlton, 1994). The parahippocampal cortex has been identified as important for declarative memory (Squire & Shimamura, 1996; Zola-Morgan et al., 1989; Miller, Lai, & Munoz, 1998). Furthermore, recent research has implicated that damage to the parahippocampal area of the brain can directly contribute to degenerative processes, such as Alzheimer's, a disease known for its severe memory impairment (Pantel, Kratz, Essig, & Schröder, 2003; Scharfman, Witter, & Schwarcz, 2000) and that parahippocampal atrophy might have potential as a predictor of Alzheimer's disease (Echávarri et al., 2011).

The rhinal cortex, which is an area of the temporal lobe just lateral to the hippocampus and amygdala, can be further divided into the entorhinal cortex and perirhinal cortices. The rhinal cortex is purported to be involved in recognition processes. Severe recognition memory deficits result from damage restricted to the perirhinal cortex (Burwell, Witter, & Amaral, 1995; Buffalo et al., 1998; Squire, 1992). Recent studies made possible by improved developments in genetic and physiological techniques have allowed researchers to explore the connectivity of the entorhinal cortex. Specifically, the separate contributions of the direct (temporoammonic) pathway from the entorhinal cortex to the CA1 subfield have been described in increasing detail, along with the indirect (trisynaptic) pathway from the entorhinal cortex to the CA1 subfield via the dentate gyrus and CA3 subfield (Bakker, Kirwan, Miller, & Stark, 2008; Nakashiba, Young, McHugh, Buhl, & Tonegawa, 2008). As the entorhinal cortex is a major source of projections to the hippocampus and the dentate gyrus, the anterograde amnesia becomes more severe when these cortical regions are damaged as well (Zola-Morgan & Squire, 1993; Zola, 1998; Zola-Morgan et al, 1989; Graham et al, 2000; Nadel, 1994).

1.4.4 Amygdala

The role of the amygdala in memory function has been shown to be restricted to nondeclarative memory, such as, the encoding of emotionally arousing stimuli and memory (Markowitsch, 2000; Miller et al, 1998; LaBar & LeDoux, 2003). It is located within the medial temporal lobe and is connected via a multitude of neurons and synapses to several important brain centres, including the neocortex and visual cortex. The amygdala also has links to the hypothalamus, and autonomic and hormonal outputs.

It is generally accepted that the connections between the amygdala and the frontal cortical regions of the brain are involved in regulating emotion and in directing emotion-related behaviours. It subserves a basic associative learning function in emotional memory and memory for emotional responses. The amygdala also serves as a potent aide for the long-term retention of emotional events and has been implicated in playing a potentially central role in the development of conditioned fear and influencing the behavioural response to a neutral stimulus from previous experience (Davis, 2006; LeDoux, 2000; McGaugh et al., 1993). It is currently generally thought that the essential plasticity supporting the fear response develops directly in the amygdala. Recent fMRI evidence has reported gender differences in the lateralisation of amygdala function for emotionally influenced memory (Cahill et al., 2004). There has been no research to date that supports its involvement in declarative memory or amnesia (Butters, 1984; Zola-Morgan et al., 1989).

1.4.5 Diencephalic structures

Studies concerning the interaction and participation of diencephalic structures in memory processes have been predominated by two topics. Firstly, which diencephalic structures participate in memory specifically, and secondly, what roles those structures play in memory processes (Zola-Morgan & Squire, 1993). Lesions within the diencephalon have been shown to disrupt memory function, which is not surprising given that substructures of the diencephalon receive afferent fibres from the medial temporal lobe (e.g., hippocampus and entorhinal cortex). Research has implicated two main areas within the diencephalon that are particularly involved in declarative memory: the anterior and dorsomedial nuclei of the thalamus and the mamillary bodies in the hypothalamus (Dusoir, Kapur, Byrnes, McKinstry, & Hoare, 1990). The thalamus and mammillary bodies, and their connecting tracts: mammillothalamic tract and ventroamygdalofugal pathway, have been found to play a significant role in episodic memory (Butters & Stuss, 1989; Kopelman, 1995; Mayes, Meudell, Mann, & Pickering, 1988; Squire, 1992; Squire & Zola-Morgan, 1988; Tranel & Damasio, 1995). These areas are purported to be specifically linked to the formation of long-term declarative memories. Lesions in these areas, as seen in Wernicke-Korsokoff Syndrome and other forms of diencephalic amnesia, can result in severe anterograde amnesia (Aggleton & Saunders, 1997). One of the main problems with research on the anatomic localisation in diencephalic amnesia has been the tendency of researchers to focus on profound memory loss following lesions of the diencephalon, to the exclusion of lesions resulting in mild or moderate memory loss (Zola-Morgan & Squire, 1993).

1.4.6 Prefrontal cortex

Lesions of the frontal lobes may also result in reduced memory functioning but not amnesias (Squire & Butters, 1992). The prefrontal cortex has been proposed to subserve what was initially termed short-term memory and, in later elaborations, working memory (Goldman-Rakic, 1995; Fuster, 2008). Working memory function is generally considered to be a subsection of declarative memory. Eichenbaum (2004) reports lesion studies indicating deficits in working memory for problem solving and planning of behaviour. The prefrontal cortex has also been identified as broadly important for processes and strategies involved in monitoring, organising, and using memory (Busch et al., 2005; Squire, 2009b).

Preliminary research has suggested that different regions of the frontal lobes may be involved in different aspects of memory functioning. Studies have identified different regions of the prefrontal cortex as playing a crucial role during remote memory recall (Frankland & Bontempi, 2005). Of specific interest, has been the role of the dorsolateral prefrontal cortex (DLPFC) in episodic memory (Kapur et al., 1995); and the recent findings that retrieval is mediated within the right side of the DLPFC, while encoding is mediated on the left (Kelley et al., 1998). The DLPFC may also be involved in recalling frequency and recency; that is, how often an event has happened and how long ago an event occurred (Anderson, Damasio & Tranel, 1993; Smith & Milner, 1988). More recently, studies have also implicated the prefrontal cortex in priming and other nondeclarative memory processes (Fletcher & Henson, 2001), in addition to a role in prospective memory (McDaniel & Einstein, 2007).

1.4.7 Striatum

The striatum, comprised of the caudate nucleus and the putamen, has been shown to be vital in nondeclarative memory formation of skills and habits (Prado-Alcala et al, 2003; Mishkin et al., 1984). This memory function, known as procedural memory, centres around memory of behaviours. Lesions of the striatum have been shown to affect procedural learning and, more recently, working memory function (Lewis, Dove, Robbins, Barker & Owen, 2004). A key feature of the structure of the striatum is its connectivity. It takes in highly processed sensory information and sends out signals involved in motor responses. Pathological disruption to striatal function causes significant impairment in memory function, in such disease processes as, Huntington's (Brandt, Shpritz, Munro, Marsh & Rosenblatt, 2005) and Parkinson's diseases (Knowlton, Mangels, & Squire, 1996; Lewis et al., 2004).

1.4.8 Cerebellum

The cerebellum is the main component in a subsystem that contributes to the execution of movement details, and to the acquisition of conditioned reflexes and body adjustments to changing environmental inputs. It receives input from restricted sensory and motor areas within the cortex and has links to the brain stem, spinal cord and thalamus (Desmond & Fiez, 1998). The cerebellum is thought to be involved in the acquisition and retrieval of nondeclarative memory function, which includes basic associative learning related to skeletal musculature; although knowledge of the precise role of these areas in memory functioning remains limited (Desgranges et al., 1998; Squire, Knowlton, & Musen, 1993; Thompson, 1986; Tranel & Damasio, 1995). Aspects of cerebellar function have also been implicated in procedural and episodic memory. This system is relatively independent and damage to these areas appears to have no effect on the functioning of the declarative memory system of the medial temporal lobe (Cohen & Squire, 1980). Damage to the cerebellum has also been reported to give rise to cognitive disturbances and cerebellar lesions inhibit the learning and acquisition of new complex skills (Appollonio, Grafman, Schwartz, Massaquoi & Hallett, 1993).

1.4.9 Parietal cortex

The parietal lobe plays an important role in tertiary processing, integrating visual and auditory sensory input. Although inputs of the various sensory systems are often reported as being independent, we experience sensory events as a single perceptual experience. The ability of the human mind to recognize concurrent sensory signals as a single percept is known as ‘cross-modal matching’. This occurs with any combination of visual and auditory stimuli; in each case the ‘matching’ is purported to take place in the areas of the tertiary cortical regions where the inputs overlap (Jones & Connolly, 1970). The majority of this process of analysis is carried out by the posterior parietal cortex, which has also been reported to be heavily involved with visuospatial processing. The ‘two-streams hypothesis’, initially put forward by Mishkin and Ungerleider (1982) is a popular, but debated, model of visual processing. They proposed that as information leaves the occipital cortex, it diverges into two ‘streams’: the ventral stream, also known as the ‘what pathway’, which terminates in the temporal lobe; and the dorsal stream, also known as the ‘where pathway’ that terminates in the parietal lobe. The dorsal stream has since been revised to be referred to as the ‘how’ stream (i.e., as in vision for action) following an important review in 1992 by Goodale and Milner. More recently, however, research has provided limited support for a

dissociation between perception and action processing of visual information in the posterior parietal cortex and instead suggested the possible existence of a more unitary processing stream (Cardoso-Leite & Gorea, 2010).

The role of the parietal cortex in memory function is not as well documented as the medial temporal lobe. Modality specific areas within the parietal cortex are argued to subserve working memory function (Squire & Zola-Morgan, 1991). For example, the lateral intraparietal area is actively engaged in visual working memory. The hypothesis of modality specific areas involved in working memory processes has also been supported by clinical observation that there are distinct auditory and visual working memory deficits in patients with schizophrenia (Huguelet, Zanello & Nicastro, 2000).

Functional neuroimaging studies, mostly using fMRI, have clearly showed activity in the posterior parietal cortex during memory retrieval (Todd & Marois, 2004). However, it has been argued that the observed pattern of parietal activation during memory retrieval reflects the region's role in directing attention during perception (Cabeza, Ciaramelli, Olson & Moscovitch, 2008). More research is needed to understand the interacting roles of attention and memory in the superior and inferior parietal cortices.

1.5 Neurophysiological processes of memory

Contemporary neuroscience philosophy embraces the neuronal doctrine, according to which, memory processes can be understood on the basis of cerebral neurophysiological processes. Just as no single brain structure accounts for all learning; so too, no single cellular mechanism accounts for all memory function. Although considerable debate continues around the molecular processes of memory, research points to changes in neurotransmitter release from neurons, fluctuations in hormone levels, and cortical protein synthesis (Kandel, 2004).

For the encoding, storage and retrieval of memories to occur, information must leave a memory trace. This has been postulated to result from some type of biochemical modification. Consolidation of information into a memory trace has been proposed to result from alterations in the strength of synapses. Long term potentiation (LTP) is a form of synaptic plasticity that has been shown to lead to long-term physiological changes in the human brain and is thought to serve as a cellular basis of memory. LTP refers to the long-term increase of the N-methyl-D-aspartate (NMDA effectiveness of synaptic transmission that can result from extended stimulation of a pre-synaptic cell

(Bliss & Lomo, 1973). LTP has been demonstrated in the CA3 and CA1 areas of the hippocampus, the dentate gyri and areas of the neocortex, all of which have been shown to be involved in normal memory function (Shepherd, 1994). Changes in N-methyl-D-aspartate (NMDA) receptor activity are believed to underlie LTP. NMDA receptors are the predominant molecular tool for controlling synaptic plasticity and, also, memory function (Panatier et al., 2006). NMDA receptor activation regulates local synaptic protein synthesis required for long-term changes in synaptic strength. Calcium flux through NMDA receptors is thought be crucial in synaptic plasticity, a cellular mechanism for learning and memory (Li & Tsien, 2009). Hippocampal NMDA receptors, in particular, have been comprehensively studied because the importance of this region in memory has been well established. NMDA receptors in the hippocampus have been demonstrated to be necessary for spatial memory function (Tsien, Huerta & Tonegawa, 1996).

The identification of differing processes for both STM and LTM have been proposed. STM involves modification of existing proteins via a second messenger, including closing calcium channels and increasing membrane potential. In contrast, LTM involves creation of new proteins, which produce more enduring channel closures and membrane potential changes (Suzuki & Eichenbaum, 2000).

1.6 Summary of neuroanatomical and neurophysiological memory processes

From reviewing the progression of the role of specific neurological structures that contribute to memory processes, the literature indicates that essential structures contributing to memory are: (1) temporolimbic structures in the medial temporal lobe, including the hippocampus, parahippocampal cortex, rhinal cortex and amygdala; (2) diencephalic structures, such as the anterior and dorsomedial nuclei of the thalamus and the mamillary bodies in the hypothalamus; (3) striatum; prefrontal cortex; and (4) to a lesser extent, the cerebellum and parietal cortex. Neurophysiological models of long-term memory and transmitter systems associated with memory, indicate that long term potentiation, through NDMA receptor activity, serves as a cellular basis of memory. The combination of increased understanding of working memory systems and development of methods to assess the neurophysiological aspects of memory processing provide an opportunity to evaluate memory functioning as a part of clinical assessment.

1.7 Assessment of memory

The importance of accurately evaluating adult memory functioning has long been acknowledged as an essential and integral part in the framework of any comprehensive neuropsychological examination (Erikson & Scott, 1977). Assessment of memory can assist in the identification of any deficits in this vital cognitive function, diagnosis of an underlying disorder or cause of memory problems, measurement of the severity and extent of dysfunction, contribute to treatment and management, help determine whether a deficit is organic or functional in origin, monitor changes in functioning over time, and help gauge the success of rehabilitation strategies (Mayes, 1986; Eslinger, 2002; Delis, 1989; Lezak, et al, 2004; Gfeller, Meldrum & Jacobi, 1995; Squire, 1986). Detailed knowledge of the integrity of memory functioning also contributes to decisions regarding surgery options (Butters, Delis & Lucas, 1995). The evaluation of memory for research purposes is important for establishing the neuropsychological profiles of specific clinical populations with varied neuropathological conditions, and also for appraising theoretical conceptualisations of memory (Howieson & Lezak, 2004; Wilson, 2004). This longstanding recognition of the necessity of memory assessment has led to a proliferation of specific assessment approaches and tools. Over time, the tools for the assessment of memory have been developed, modified and refined; however, the majority of assessments have not been based on conceptual models, but rather on clinical values that have then been either confirmed or refuted (Butters, Delis & Lucas, 1995).

The type of memory assessment varies with the age and condition of the individual being assessed. A functional memory assessment generally tests the individual's ability to recall information and makes an assessment of their memory skills related to their home, vocational and social environments (Lezak, et al, 2004). Of all the higher cognitive functions, memory function appears to be particularly vulnerable to the deleterious effects of a myriad of brain dysfunctions (Mayes, 1995). As a general rule, performance on memory tests is poorer after an individual has sustained any injury to the brain. Impaired memory function is a frequently observed occurrence among patients with wide ranging diagnoses such as: encephalitis, stroke, dementia, cerebral tumours, acquired brain injury, anoxia and epilepsy (Chelune & Bornstein, 1988; Baddeley, 1995; Kapur, 1988; Loring & Papanicolaou, 1987; Squire, 1986; Squire & Shimamura, 1996; Tulving, 2002).

1.7.1 Issues in the clinical assessment of memory

Memory problems are a common sequel of neurological trauma and disease and are often reported in association with affective disorders (Eichenbaum, 2002). Among patients referred for neuropsychological assessment, disturbed memory function is one of the most common presenting complaints (Caplan & Caffery, 1996). Disruption of memory function, such that occurs in amnesic disorders and dementia syndromes can have devastating effects given the importance of learning and memory in most aspects of life (Squire, 2009a; Troester, 1998). Despite this, many memory problems do not represent a complete loss of memory (Squire, 1987). Indeed, in some conditions memory for certain kinds of information can be intact, while memory for other types of information can be impaired (Mayes, 1995; Wilson, 2004). In addition, memory difficulties can occur secondary to impairment in other cognitive functions, such as attention and concentration (Reid & Kelly, 1993). There are similarities in memory functioning between some forms of amnesia and dementia; however differences can be observed, consequently, memory assessment procedures need to be adequate to address these different patterns of memory impairment (Butters, Delis, & Lucas, 1995; Papanicolaou, 2006).

Assessment of memory function also needs to allow for the memory changes that occur as a normal part of aging. Indeed, the pattern of memory decline in elderly people has been shown to be variable (Bornstein & Chelune, 1989; Fahle & Daum, 1997). Petersen, Smith, Kokmen, Ivnik and Tangalos (1992) demonstrated that while acquisition performance decreased with age, delayed recall remained stable across age when the amount of material initially learned was controlled for. Therefore, normative data for geriatric populations is necessary to control for the observable pattern of normal decline with increasing age (Ivnik et al., 1992). Clearly the clinical assessment of memory in elderly populations requires this factor to be taken into consideration.

The neuroanatomy of memory is generally modality specific, at least at a hemispheric level, that is, the left hemisphere mediates verbal memory and the right hemisphere mediates visual memory (Parkin, 1997; Frisk & Milner, 1990; Lee et al, 2002). Therefore, we might expect reasonably discrete amnesic syndromes for specific hemispheric lesions; however, this only holds true with relatively circumscribed lesions (Bornstein, 1982; Chelune & Bornstein, 1988; Lee, Loring & Thompson, 1989). Lesions in the left temporal lobe have consistently demonstrated deficits in learning for verbal information. Conversely, lesions in the right temporal lobe have been shown to

result in visual memory deficits, although the findings are not as uniform (Chelune & Bornstein, 1988; Squire & Butters, 1992; Squire, 1986; Jones-Gotman, 1986; Naugle, Chelune, Cheek, Luders & Awad, 1993). The inconsistencies in the findings with regard to visual memory are thought to be partly explained by methodological issues and the challenges in developing a true measure of non-verbal memory. However, recent functional magnetic resonance imaging (fMRI) studies have indicated bilateral representation of visual memory function. Using patients with unilateral mesial temporal lobe epilepsy (MTLE) and matched controls it has been demonstrated that a complex network of connections in the parietal, temporal, and frontal lobes of the left hemisphere were activated in verbal memory processing. Results for visual memory indicated a more diffuse and bilateral representation both in controls and in patients with MTLE (Figueiredo et al., 2008; Alessio et al., 2011).

Additionally, memory function can be impacted by lesions anywhere in the brain. For example, language difficulties can impact verbal memory and parietal lobe lesions can impact visual memory (Mayes, 1986). Furthermore, most visual memory tests include materials that can be verbalized to a certain extent. The problem is that so many subsets of factors exist within the memory process that assessment of memory must be careful in grouping individuals based solely on a simplistic diagnoses (Jurden, Franzen, Callahan & Ledbetter, 1996; Russell, 1975). The clinical assessment of memory must take a multifaceted view of a diverse set of factors that help form the individual memory of each human being (Eslinger, 2002).

1.7.2 Approaches to the clinical assessment of memory

Traditionally, approaches to the assessment of memory function and test development have largely focused on have covered short-term and long-term memory techniques, and declarative memory, including episodic and semantic memory (Lezak, et al, 2004). The majority of memory testing approaches have focused on episodic memory, presumably because firstly, episodic memory function is most often affected by neurological lesions and secondly, it is thought to be particularly relevant for everyday adaptive function (Mayes, 1986).

Declarative memory involves conscious recollection generated by direct efforts to access memories. Explicit tests of memory involve direct inquiries that ask the subjects to refer to a specific event of learning or a specific fact in their knowledge (Mayes, 1995). The full range of declarative memory tests includes large variety of direct

measures of recall or recognition of word or picture lists, paired associates, story recall, and include most of the common tests of memory that are performed poorly by amnesic patients (Butters, Delis & Lucas, 1995; Loring & Papanicolaou, 1987).

Implicit memory involves unconscious changes in performance of a task as influenced by some previous experience. Implicit tests of memory involve indirect measures such as changes in the speed of performance or biases in choices made during performance of a task that can be solved with the information at hand (Squire, 1986). Examples of implicit memory tests include the full variety of assessments of motor, perceptual, and cognitive skills, habits, conditioning, and repetition priming at which amnesic patients usually succeed (Mayes, 1986). Notably none of these tests requires the subjects to be aware of their memory, or to 'remember', a specific event or fact. (Delis, 1989; Squire & Shimamura, 1996).

1.7.2.1 Assessment of non-verbal memory

The assessment of declarative memory has been further refined to include tasks that reflect the distinction of verbal and non-verbal memory abilities (Smith, Malec, & Ivnik, 1992). The majority of tests, normative data and research into memory functioning have focused predominately on verbal memory function, specifically short-term verbal memory functioning. The assessment of non-verbal memory function, although vital, in neuropsychological assessment, has received comparatively little attention (Heilbronner, 1992). There is a well reported dissociation between verbal and non-verbal memory performance in various clinical groups, so any comprehensive assessment of memory function needs to address both verbal and non-verbal abilities (Parkin, 1997; Lee et al, 2002).

Measures of non-verbal memory have typically included: tests of immediate visual span; serial recall of visuospatial information; recognition of non-verbal information, such as, faces; recall, recognition and reproduction of abstract designs (Moye, 1997; Lezak et al, 2004). Immediate visual span measures, such as the Corsi blocks (Milner, 1971) and the spatial span subtest (Wechsler, 1987), involve the repetition of a particular sequence of taps in the same manner as the sequence was presented to evaluate immediate memory for visuospatial information.

Among the most common tests of nonverbal memory used by neuropsychologists are measures of figural reproduction. These tests generally involve a copy or immediate

reproduction of an abstract geometric figure or figures, followed by a delayed recall after a predetermined time period. The most common memory tests for figural reproduction include: Rey Complex Figure Test (Rey, 1964; Meyers & Meyers, 1995), Benton Visual Retention Test (Benton, 1992), Brief Visuospatial Memory Test (Benedict & Groninger, 1995) and the Visual Reproduction subtest of the Wechsler Memory Scale and its revisions (Wechsler, 1945; Wechsler 1987).

One of the earliest visual memory measures was the Rey-Osterrieth Complex Figure Test (ROCFT: Rey, 1964) which required an individual to recall and reproduce a complex design after a brief presentation. However, the ROCFT was designed to assess the integrity of different functions which, although including visual memory, also evaluated perception, visuospatial abilities, attention, planning and executive functions. Benton developed the Benton Visual Retention Test (BVRT) in 1974. The BVRT assessed only immediate and short-term memory, and did not allow for the assessment of long term memory function. The BVRT was widely used and included three alternate test forms. However, the evidence supporting the inter-form reliability was generally weak.

In 1995, Benedict and Groninger published the Brief Visuospatial Memory Test (BVMT). The test was produced with six alternate forms to address the lack of a multi-form visuospatial memory test with established equivalence. The test required subjects to immediately draw visually presented abstract designs after a 10-second exposure, followed by a 25-minute delayed recall. The authors reported satisfactory equivalence of the six alternate forms, but noted that the incorporation of learning trials and delayed recognition measures would improve the clinical utility of the test. Arguably one of the most well-known and well used measures of nonverbal memory has been the Visual Reproduction subtest from the Wechsler Memory Scale, which will be covered in greater detail in section 1.9.

Criticisms of the use of figural reproduction tests to assess visual memory have identified the need for perceptual and constructional deficits to be ruled out as confounds in visual memory recall (Caplan & Caffery, 1996; Heilbronner, 1992). All of the tests covered above rely on perception and construction skills, consequently intact visual memory abilities can be masked by praxis impairment or perceptual disturbance and this can lead to misdiagnosis of the extent and nature of non-verbal memory impairments (Smith et al, 1992; Larrabee & Curtiss, 1995; Haut, Weber, Wilheim,

Keefover & Rankin, 1994; Loring, 1989; Moye, 1997; Bowden, Ritter, Carstairs, Shores, Pead, Greeley et al, 2001; Fastenau, 1996; Gfeller et al., 1995). A limited number of standardised measures have included recognition procedures to provide information on the relative contributions of encoding, storage and retrieval on visual memory performance (Glosser et al, 1989; Meyers & Meyers, 1995).

One of the other main difficulties in using a figural reproduction tests to assess visual memory has been the unwanted potential for verbally-mediated memory process involvement (Heilbronner, 1992; Moye, 1997). The difficulty of developing a visual memory task that is not confounded by verbal mnemonics has resulted in the artefact of the test not exclusively addressing non-verbal memory functions. The fact that the majority of scoring criteria are entirely verbal has exacerbated the potential for verbal encoding. Nevertheless, in the absence of a 'perfect' measure of non-verbal memory, figural reproduction tests have been used extensively (Heilbronner, 1992). In a more recent attempt to examine the construct of nonverbal memory, as assessed by figural reproduction tests no significant difference was demonstrated in performance between surgery candidates with right or left temporal lobe epilepsy (Barr et al., 1997). These findings once again highlighted the major inadequacies of contemporary figural reproduction tests and introduced uncertainty about the processes of nonverbal memory being exclusively mediated by right hemisphere functions.

A review of these select memory assessment tools has demonstrated the complex nature of memory assessment, particularly for non-verbal memory processes, and emphasised the need for development of a more sensitive and comprehensive memory assessment tool for use in a range of clinical scenarios. In the next section the development of one the most commonly used memory test batteries in clinical settings, the Wechsler Memory Scales, is discussed.

1.8 Wechsler Memory Scales

1.8.1 The Wechsler Memory Scale

In 1945, David Wechsler published the Wechsler Memory Scale (WMS) with the aim of creating a "rapid, simple, and practical" measure in the clinical assessment of memory and memory disorders (Wechsler, 1945, p. 3). At the time, its development represented a breakthrough in the clinical assessment of memory and it was one of the earliest attempts at a standardised test battery. It quickly became arguably one of the most commonly used memory test batteries (Russell, 1975) and its subtests became the focus

of a substantive component of the early research into clinical memory assessment (Heilbronner, 1992). In spite of its extensive use in clinical practice and neuropsychological research, the WMS received widespread criticism on both theoretical and psychometric grounds, with regards to its normative data, reliability, factor structure, construct validity, scoring criteria, conceptual aspects and clinical utility (Erikson & Scott, 1977; Prigatano, 1978).

1.8.1.1 Structure and content

The WMS consisted of seven subtests: Personal and Current Information, Orientation, Mental Control, Logical Memory (LM), Memory Span, Visual Reproduction (VR), and Associate Learning; the origins of which were argued to be strongly influenced by Wechsler's military psychological testing experiences (Boake, 2002). Critics identified areas of weakness in the composition of the WMS including: a high reliance on short-term verbal memory, with only one test of visual memory; testing only free recall, with no recognition or cued recall; no focus on the role of learning; and the inclusion of tests that addressed functions other than memory, such as attention and concentration (Chelune & Bornstein, 1988; Delis, 1989; Erikson & Scott, 1977; Prigatano, 1978).

For the purposes of repeated memory assessments an alternative form (Form II) of the WMS was also developed for subsequent assessments. Form II was closely matched to Form I with the same seven subtests, but with changes to subtest content. Wechsler reported a satisfactory degree of equivalence between the two forms (1945). However, Bloom (1959) demonstrated that the two forms were not entirely comparable, and that in Form I the Associate Learning test was easier and that the Visual Reproduction subtest was easier on Form II. No empirical data were provided to support the claim that the WMS was a reliable measure. No information regarding the test-retest reliability of the WMS or the internal consistency of the subtests of either Form I or Form II were reported in the WMS Manual (Prigatano, 1978).

1.8.1.2 Administration, scoring and normative data

Administration of the WMS was generally straightforward with adequate test instructions, user-friendly materials and guidelines provided for scoring. However, the scoring procedures for two subtests, Visual Reproduction and Logical Memory, were specifically criticised for being vague, insufficient and subjective (Loring & Papanicolaou, 1987), resulting in poor inter-scorer reliability.

Subtests were scored on the basis of single point for a correct response and no points for an incorrect response. Subtest scores were then added together with an age correction factor. A global 'memory quotient' score (MQ) was then derived via summation of the scores of verbal and visual memory, including the scores on the measures of orientation, attention and concentration (Wechsler, 1945). This conceptualisation of human memory as a unitary construct was at odds with research findings (Lezak et al., 2004) and limited the sensitivity of the WMS as a diagnostic instrument capable of identifying discrete memory dysfunction (Parkin & Leng, 1993). Furthermore, the dichotic factor structure of the WMS was later determined to be very weak (Larrabee, Kane & Schuck, 1983), although Wechsler's assumption that the MQ would be directly comparable with performance on intellectual ability measures was summarily supported (Ryan, Rosenberg & Heilbronner, 1984).

Standardisation of the WMS was reported to be based on a normative sample of 200 'normal' subjects aged 25-50 years of age. However the manual reported data for two groups of participants: 50 aged 20-29 years and 48 aged 40-49 years (Wechsler, 1945). Either way, the standardization sample was universally described as inadequate and disparaged for failing to report the gender percentage of the sample or any performance differences between sexes (Loring & Papanicolaou, 1987). The truncated age range of the normative sample precluded its applicability and clinical utility in aging populations, in spite of the elderly frequently requiring memory assessment (Prigatano, 1978).

Various revisions to the administration were proposed in response to the well-recognised limitations of the WMS (Mitrushina, Boone, Razani, & D'Elia, 2005). Endeavours to improve the administration of the WMS were proposed firstly by Russell (1975) and later by Milberg, Hebben and Kaplan (1986). These modifications to the original WMS endeavoured to broaden the recall options to better evaluate both STM and LTM. Both versions provided delayed recall trials for the prose memory test, Logical Memory (LM), the graphomotor memory test, Visual Reproduction (VR), and the word-pair learning task, Associate Learning. Russell (1975) also provided enhanced scoring procedures for the LM subtest and included recognition, copy and perceptual match trials for the VR subtest. Russell (1975) also introduced the concept of calculating 'saving scores' (i.e., percent retention scores) for both the LM and VR subtests. Milberg et al. (1986) also included copy and perceptual match trials to cater for sensorimotor deficits on the VR subtest.

Although the widening of the recall options was welcomed by clinicians, the lack of clinical data restricted the usefulness of these revisions. In addition to modifications of the administration and scoring, improved collections of normative data were also established in response to the identified inadequacy of the standardization sample. Normative data were published for adolescents and older adults of various population groups (See Mitrushina et. al., 2005 for a detailed review).

1.8.1.3 Clinical utility of the WMS

The general consensus reached about the WMS was that it was a sensitive test of short-term verbal memory, limiting its clinical utility to identifying amnestic disturbances associated with left temporal lobe impairment and dysfunction in its medial hippocampal connections (Prigatano, 1978). Recommendations for improvement included complete re-standardisation. Furthermore, the importance of an Australian standardization sample to enhance clinical utility in an Australian population was put forward by Ivison (1977). Despite its numerous limitations, the WMS was widely used in clinical practice and generated a substantial body of research.

1.8.2 The Wechsler Memory Scale-Revised (WMS-R)

Before his death in 1981, Wechsler initiated an extensive revision and re-standardisation of the WMS. As a result, after four decades of widespread use and spirited debate, the WMS was superseded in 1987 by its formal revision: the Wechsler Memory Scale-Revised (WMS-R: Wechsler, 1987). The revised version of the WMS aimed to better assist the clinical evaluation of memory functions and memory disorders (Wechsler, 1987). Major improvements were inclusion of measures of delayed recall, more specific administration guidelines, greater ranges in score, and more detailed scoring criteria. The revision was generally considered to be a 'purer' measure of memory that yielded results that were more consistent with other memory instruments than the original version (Petersen et al., 1992).

1.8.2.1 Structure and content

The Logical Memory, Visual Reproduction and Associate Learning subtests were identified by Butters (1984) as the most clinically useful measures of memory on the WMS and all three were retained as part of the WMS-R, with Associate Learning, renamed as Verbal Paired Associates (VePA). The WMS-R also retained three other subtests that originated in the WMS: Mental Control; Information and Orientation, which combined the Information and Orientation subtests from the WMS; and Memory

Span, renamed as Digit Span. Three new measures were included in the WMS-R: Figural Memory, Visual Paired Associates and Visual Memory Span, which brought the number of subtests in the WMS-R to nine in total.

Story A was retained in the LM subtest story with some slight alterations to the content. A new story was developed to replace story B in an attempt to achieve better equivalency between stories A and B. In the VR subtest card A and card B were retained with two newly developed designs replacing cards C and D. Two of the easier word pairs were dropped from the VePA subtest in an attempt to shorten it. However, A further three trials were added so that learning could be examined in greater detail. Procedures for delayed recall were also developed for the LM, VR and VePA measures. Minor changes to the other retained subtests included removal of the speed bonus on the Mental Control subtest and the addition of more trials of a shorter digit sequence on the Digit Span subtest.

The original WMS received significant criticisms for principally being a test of immediate verbal memory, with only one subtest purporting to assess non-verbal memory. To provide a more comprehensive assessment of visual memory ability three non-verbal memory subtests were developed: Figural Memory (FM), Visual Memory Span (VMS) and Visual Paired Associates (ViPA). The FM subtest was reported to measure memory for figural material and required an individual to identify target patterns from a set of designs. The ViPA subtest was intended to be an analogue to the VePA subtest and required the individual to recall the association between line drawings and colours. The VMS was conceived as an analogue to the verbal Digit Span subtest and involved copying an increasing sequence of squares being tapped in a predetermined order (Loring, 1989).

Regrettably, the inclusion of additional measures to address non-verbal functioning did not achieve the overall aim of assuaging the predominantly verbal bias of the WMS (Lezak et al, 2004). The FM subtest was identified as relying on higher order processes, such as visual attention, rather than visual memory and was further criticised for having no procedure to assess long-term retention of the material (Loring, 1989). The VMS subtest was also found to load on both memory and attentional factors (Bornstein & Chelune, 1988). Factor analysis revealed that the ViPA grouped together more with verbal than non-verbal measures, which was suggested to be due to the relative ease

with which the task could be verbally encoded (Loring, Lee, Martin & Meador, 1989; Wong & Gilpin, 1993).

Further shortcomings of the structure and content of the WMS-R included: a procedure to assess cued recall in the LM subtest but not the VR subtest, making specific performances between the VR and LM subtests difficult to compare; and the absence of an alternative form for subsequent administration (Lezak et al., 2004; Loring, 1989). The inclusion of only limited recognition procedures resulted in difficulty differentiating the role of recall and recognition in poor performances (Butters et al., 1988; Troester et al., 1993). However, several authors have subsequently developed their own recognition options for the LM and VR subtests, improving the clinical utility (Fastenau, 1996; Gass, 1995).

1.8.2.2 Administration, scoring and normative data

In light of the criticisms of the original WMS, five composite indices: Attention/Concentration, General Memory, Verbal Memory, Visual Memory, and Delayed Recall, replaced the single MQ score from the WMS (Loring, 1989). Initial results of confirmatory factor analytic studies were varied. Jurden, Franzen, Callahan, and Ledbetter (1996) demonstrated satisfactory factorial equivalence of the WMS-R between the original standardization sample and substance abusing inpatients, who were reported to have diffuse neuropsychological pathology. However, many authors reported that the factor structure of the WMS-R co-varied according to the population group and with age and years of education (Bornstein & Chelune, 1988; Loring et al., 1989; Wilhelm & Johnstone, 1995). Further significant modifications incorporated into the WMS-R included: more specific administration guidelines; a full revision of scoring procedures for several subtests, including more detailed scoring criteria; a greater range in possible scores; and normative data for different age levels from 16 to 74 years (Reid & Kelly, 1993; Williams et al., 1998).

Improved scoring criteria were also included for the LM and VR subtests, which were designed with the aim of minimising scoring complexity and maximising scoring accuracy (Wechsler, 1987). The development and inclusion of detailed examples of the stories and reproductions of the drawings further aided in clarifying the scoring principles. The WMS-R manual reported an inter-scorer reliability coefficient of .97 for VR using a healthy normative sample. Woloszyn, Grob-Murphy, Wetzel, and Fisher (1993) demonstrated highly satisfactory inter-rater reliability for two of the WMS-R

subtests: LM and VR in a mixed clinical population, which was consistent with the findings in Wechsler's normative sample.

It is a widely recognised fact that psychological tests assessing cognitive function require periodic revision of their content. Revisions of the original WMS have maintained its wide acceptance and also addressed the recognised importance of updating test standardisation (Flynn, 1998). However, to procure an accurate interpretation of test results, the requisition of reliable normative data has also been identified (Mitrushina et al., 2005; Spreen & Strauss, 1998). The WMS-R normative sample size of 316, although reasonable in comparison with the Wechsler Adult Intelligence Scale-Revised (WAIS-R), had several limitations. Firstly, the norms for the age ranges 18 to 19 and 25 to 34 years-of-age were statistically estimated, by interpolation from adjacent age groups, as opposed to being derived empirically. This method of statistical estimation has been shown to produce inaccurate normative data (Mittenberg, Burton, Darrow & Thompson, 1992) and was further criticised for being based on a relatively small standardization sample (Loring, 1989; Mitrushina et al., 2005). Secondly, details of individual normative data were not available for several of the subtests. Percentile norms were provided for the immediate and delayed components of the LM and VR subtests and for the forward and backward trials of the Digit Span and Visual Memory Span subtests. However, the lack of easily comparable scaled scores between measures precluded the clinical utility of depicting various profiles of clinical populations (Wilhelm & Johnstone, 1995). In addition, the WMS-R normative data were provided only to the age of 74, limiting the evaluation of memory problems as a result of degenerative disorders in later life. This inadequacy was addressed in 1992 by Ivnik et al., who published normative data for individuals aged 56 to 94 years on the WMS-R subtests. Subsequent normative studies were conducted to address the identified inadequacies of the original standardisation sample, but were predominately based on North American populations (Mitrushina et al., 2005). The importance of establishing local normative data to accurately assess memory functioning within specific populations has also been emphasised (Ivnik et al., 1992; Levin et. al., 1987; Mittenberg et al., 1992; Walker, Batchelor, & Shores, 2009). Indeed Holdnack, Lissner, Bowden and McCarthy (2004) reported that there have been concerns surrounding the uptake of the Wechsler Memory Scales in Australia due to the absence of local norms.

The Macquarie University Neuropsychological Normative Study (MUNNS) was conducted in the late 1990's with the aim of providing local normative data for

commonly used neuropsychological assessment measures, including the WMS-R (Carstairs & Shores, 2000). The published normative data for an Australian population aged 18 to 34 years provided an advantage to utilising the WMS-R for this cohort in Australia. The MUNNS normative data were published to provide national standards against which the test performances of brain-injured Australian patients could be directly compared. Analysis of the MUNNS data suggested that US norms were not identical to Australian norms and that results needed to account for gender and demographic variables (Shores & Carstairs, 2000).

1.8.2.3 Clinical validity of the WMS-R

The wide acceptance and use of the WMS-R generated prolific research around its clinical validity, particularly within diverse clinical populations, such as subjects with: unilateral brain lesions, amnesia, Alzheimer's disease (AD), Huntington's Disease (HD), Multiple Sclerosis (MS), alcoholism, schizophrenia, TBI define and closed head injury. Encouragingly the results of this research were almost uniformly positive. Butters et al. (1988) examined the clinical validity and sensitivity of the WMS-R in the differentiation of amnesic patients from those with dementia. Sixteen amnestic patients, 20 patients with AD, 24 patients with HD and 28 normal control subjects were administered the WMS-R. The authors reported that amnestic patients could be distinguished from patients with cortical and subcortical dementias, and control subjects on the basis of the differences between the two main WMS-R Indices; General Memory and Delayed Memory.

The use of the Wechsler Memory Scale-Revised to detect and characterize memory deficits in MS was investigated by Fischer (1988). A sample of 45 patients with MS and 25 normal controls were administered the WMS-R. As a group, the patients with MS performed significantly worse compared with the normal controls on all five WMS-R indexes. The authors concluded that the WMS-R demonstrated satisfactory clinical validity and sensitivity in the detection of memory impairment in MS. They also reported that the degree of impairment was not related to demographic, disease characteristics, medication status, or mental illness, thus providing further evidence of the overall satisfactory clinical validity of the use of the WMS-R in the MS population.

Ryan and Lewis (1988) examined the performances of recently detoxified alcoholics with normal controls on the WMS-R. The alcoholic subjects performed significantly more poorly than the matched control subjects on all five WMS-R index scales,

similarly to the patients with MS. Furthermore, the performance of 20 subjects with closed head injury on the WMS-R was compared with matched controls. The subjects in the closed head injury group also performed worse than controls on all five WMS-R indices (Reid & Kelly, 1993). In addition to demonstrating the sound clinical validity of the WMS-R in their population of alcoholics, Ryan and Lewis (1988) also provided evidence of ecological validity by reporting that observed memory status was strongly associated with WMS-R index scores. This finding was supported by Reid and Kelly (1993) who reported that patients with a closed head injury who performed worse on the WMS-R also received poorer ratings on an independent assessment of everyday memory.

The application of the WMS-R to a psychiatric sample was investigated by Gold, Randolph, Carpenter, Goldberg and Weinberger (1992). They examined the performance of 45 patients with schizophrenia on the WMS-R whose results indicated memory impairment when compared with the WMS-R normative sample. The researchers concluded that the findings supported the validity of the clinical use of the WMS-R to detect memory impairment in patients with schizophrenia

The development of separate verbal and visual indexes was predicted to improve WMS-R sensitivity to lesion laterality. In 1988, Chelune and Bornstein examined the patterns of performance on the WMS-R of 115 patients with unilateral brain lesions. As hypothesised, the patients with left-hemisphere lesions experienced greater difficulty learning and retaining verbal material than comparable nonverbal/visual material, while the right-hemisphere lesion patients demonstrated the opposite pattern. The authors concluded that their results supported the clinical validity of the WMS-R as a multivariate measure of modality-specific memory functioning. However, the right and left lesion groups performed significantly differently on the WMS-R verbal subtests, but not on the visual memory subtests. Furthermore, studies by Loring et al. (1989) and Naugle et al. (1993) also demonstrated that the summary indices were inconsistent at identifying lateralized temporal lobe lesions. Interestingly, significantly poorer performances on the Verbal Index were observed in subjects following a left temporal lobectomy, but performance was not reduced on immediate or delayed procedures relative to preoperative baseline following right temporal lobectomy (Naugle et al., 1993). The low reliability of the WMS-R factor structure and the artefact of verbally mediated encoding on visual tasks were proposed as possible reasons for the factor structure failing to support the differentiation between verbal and visual memory

(Chelune & Bornstein 1988; Elwood, 1991). Suggestions for improved clinical utility included investigating performance on the individual WMS-R subtests, rather than relying on composite Index scores to accurately portray memory abilities (Wilhelm & Johnstone, 1995).

This section has provided an overview of the WMS, the WMS-R and the subtests. The Wechsler Scales are some of the most commonly used measures of memory function. This review of literature has demonstrated that the revised version has strengths in the inclusion of measures of delayed recall, improved scoring criteria, greater score ranges, and more detailed scoring criteria were also identified. However, weaknesses, including: a predominantly verbal bias, statistically estimated normative data, and poor sensitivity to lesion laterality were also identified. The importance of normative data in test result interpretation was also emphasised. Research continues to evaluate the WMS-R and its subtests to ensure that results yielded are accurate and useful in contemporary clinical practice. The following section provides a selective review of literature on the use and interpretation of the Visual Reproduction subtest from the WMS-R, which is the major focus of this current research.

1.9 Visual Reproduction

1.9.1 Design

One of the most commonly used measures of non-verbal memory in neuropsychological assessment has been the Visual Reproduction subtest of the Wechsler Memory Scale, and its revision in the WMS-R (Paniak, Murphy, Miller and Lee, 1998). Developed originally as a measure of immediate recall, the VR subtest of the WMS had two variations, but due to the absence of psychometric data for the second form, the first form was the only one used clinically (Prigatano, 1978). The design of the original VR subtest from the WMS required patients to draw from memory various geometric designs after being shown them for a brief period (Wechsler, 1945).

The Visual Reproduction subtest of the Wechsler Memory Scale did not include a delayed recall trial, with the absence of such a procedure limiting its clinical utility (Prigatano, 1978). This inadequacy was addressed when a delayed recall procedure was later developed, standardised and validated clinically by Trahan, Quintana, Willingham and Goethe (1988). Revisions to the VR subtest in the WMS-R introduced several improvements on the original, including: four designs compared to the previous three, allowing a more comprehensive examination; and also the inclusion of a delayed recall

measure. The addition of a standard delayed recall procedure enabled a measure of retention of information over time to be calculated and, as delayed recall provides a superior measure of memory function, made it a more useful measure of visual memory than the original (Larrabee & Curtiss, 1995; Lezak et al, 2004).

1.9.2 Clinical utility

The VR test was suggested to be a complex test that was quite sensitive to changes in brain function, as it placed little reliance on over-learned skills with its immediate demands for attention and novel encoding (Pliskin et al., 1996). Along with other WMS-R subtests, the Visual Reproduction subtest was shown to differentiate typical forgetting seen in an aging population compared with abnormal forgetting associated with Alzheimer's disease (Cullum, Butters, Troester & Salmon, 1990). Patients with Alzheimer's Disease were also shown to have more intrusion errors and evidenced more rapid forgetting than the control subjects, or patients with other degenerative disease processes, such as, Huntington's Disease (Butters et al., 1988; Jacobs et al., 1990; Troester et al., 1993).

1.9.3 Validity

The Visual Reproduction subtest was initially included in the WMS only as a contrast to verbal memory tasks, and was criticised for the likelihood that it did not solely address non-verbal memory functions; a criticism that was also levelled at its successor, the revision of the VR subtest in the WMS-R (Prigatano, 1978). There has been difficulty in designing an adequate measure of visual memory which minimises the contamination effect of other cognitive processes, such as: motor functioning, verbal elements, visuoperceptual and visuoconstructional abilities (Heilbronner, 1992).

In studies of factor analysis, the VR subtest has consistently loaded primarily on visual-perceptual motor ability, and only secondarily on memory, as demonstrated in five studies involving a significant number of participants. (Bornstein & Chelune, 1988; Heilbronner, Buck & Adams, 1989; Larrabee & Curtiss, 1995; Larrabee et al., 1983; Larrabee, Kane, Schuck & Francis, 1985) raising questions about the validity of this test as an adequate measure of visual memory processes. It was argued that sufficient visual perceptual and visuoconstructional skills, together with adequate motor functioning, were required for the accurate reproduction of visual designs. Therefore, the interpretation of performance was likely to be distorted by perceptual dysfunction, constructional deficits or impaired motor abilities (Gfeller et.al, 1995).

Although there is a relationship between the Visual Reproduction subtest and other cognitive processes, it still has a useful role in the assessment of visual memory. As a higher order process, memory and memory test performances will, most likely, reflect the integrity and role of lower order cognitive processes (Mayes, 1986). The role of the clinician is to ascertain to what extent poor performances on visual memory tasks are due to memory impairment, as opposed to other primary deficits. The component cognitive processes need to be identified so that they may be ruled out as a deficit affecting other areas of performance on a memory test (Squire & Shimamura, 1996).

The opportunity to utilise dual (i.e., verbal and visual) encoding as a strategy, and the fact that the scoring criteria were entirely verbal, highlighted the potential for contamination by verbally-mediated memory processes (Pliskin et al., 1996; Smith et al., 1992). Indeed, studies indicated that people with right hemisphere or left hemisphere damage both exhibited impairments on the subtest (Barr et al., 1997; Bigler et al., 1996; Chelune & Bornstein, 1988; Heilbronner, 1992; Trahan et al., 1988). Whilst tasks can be described as measuring non-verbal memory, it is likely that people use verbal skills when remembering visual information, particularly, if the task involves simple geometric figures that can be verbally encoded. To circumvent the confound of verbal encoding, Lee et al. (1989) suggested that a visual memory test should use complex stimuli that are unfamiliar and thus challenging to verbally encode; however, such complex designs may be too difficult for compromised patients to recall.

1.9.4 Modifications

Although the Visual Reproduction subtest was a commonly used measure of non-verbal memory, these limitations required addressing to improve its clinical utility (Heilbronner, 1992; Loring, 1989; Loring & Papanicolaou, 1987). Consequently, several authors proposed varying supplements and modifications to the design and administration of the Visual Reproduction subtest.

The VR subtest has been criticised because it fails to control for, or correct for, perceptual and constructional deficits. Researchers who have identified these criticisms have also attempted to expand the clinical utility of the VR subtest and address these particular limitations through the application of supplemental procedures. Published modifications have included: a recognition trial or cued recall (Fastenau, 1996; Gass, 1995; Hanger, Montague and Smith, 1991; Haut et al., 1994; Wilhelm, 1996), a matching and copy trial (Fastenau, 1996; Haut et al., 1994), a constructional skill

proportion score (Haut, Weber, Demarest, Keefover and Rankin, 1996), and the calculation of percentage of information retained and scoring of intrusion errors (Wilhelm, 1996). Clinical utility is greatly enhanced, as a copy trial allows for an examination of constructional ability and a matching trial permits investigation of perception, both of which are prerequisite for then considering recall and recognition abilities.

The incorporation of a recognition trial or cued recall has been the most widely suggested adjunct to the standard administration as it provides an intermediary measure of delayed recall. It is clinically useful to note where a partial cue sufficiently triggers recall. Hanger et al. (1991) developed a recognition trial of the VR subtest, and demonstrated that this could differentiate between a neurologically impaired group and a neurologically normal group. In 1994, Haut et al. also developed a recognition trial, but went further and developed matching and copy trials for the VR subtest. These measures were then used to compare the performance of thirteen patients with Alzheimer's disease (AD) with thirteen neurologically impaired patients and fourteen neurologically normal controls. The results showed that the subjects with AD were significantly poorer at matching and copying the VR designs in comparison to the neurological and non-neurological control groups. These results indicate subjects with AD exhibited impaired construction and visual perceptual abilities. In a further extension of their previous study, Haut et al. (1996) attempted to control and correct for the contribution of constructional skills on the Visual Reproduction subtest. In a group of patients with AD they calculated a proportion score by dividing the subtest raw scores from standard administration by the total score obtained for copying the designs. Using this method they were able to demonstrate that, independent of constructional skills, patients with AD showed greater visual memory impairment compared to control subjects. The researchers concluded that the standard administration of the VR subtest was insufficient as a measure of memory functioning in patients with AD.

In 1995, Gass developed a procedure for assessing storage and retrieval memory processes on the Wechsler Memory Scale-Revised, which was designed to distinguish deficient memory storage from compromised retrieval. A cuing technique for VR was administered to 94 psychiatric inpatients and 99 patients with brain-injury (BI). The results showed poorer performances by the BI sample than the psychiatric sample on the standard VR measures, with VR cuing technique showing the highest degree of discrimination. From these results, Gass concluded that BI adversely affected

performance on memory tests to a greater degree than mental illness. Gass also concluded that as subjects initially acquired more information than they reported during the standard free recall format, the need for a cued or recognition VR component was evident.

Fastenau (1996) supported the notion that poor recognition memory of figures is a more reliable indicator of memory dysfunction than poor recall *per se* (Squire and Shimamura, 1996). He also reported that perceptual and constructional deficits need to be ruled out as confounds in visual memory recall. As a result he developed an elaborate administration process for the VR subtest from the WMS-R, which was designed to address the perceived limitations of the standard administration in light of these assumptions. Recognition, matching, and copy trials were added as an adjunct to the standard administration of the VR subtest. Validity coefficients were reported as satisfactory for all of the new trials ($r = .60$-1.0) and reference data were provided for sample of 81 healthy adults. The importance of additional validation using the elaborated administration in clinical groups was reported.

Wilhelm (1996) proposed a supplemental scoring system for the Wechsler Memory Scale-Revised which included several modifications to the VR subtest. These revisions included the addition of a recognition task following the delay procedure and scoring of intrusion errors. The WMS-R had previously been criticised because of its failure to recognise the importance of examining delayed recall as a percentage of information initially recalled (Troester et al., 1993). Wilhelm also included a calculation of percentage of information retained from immediate recall to delayed recall. Results indicated that the percent retained measure was significantly different between clinical groups of chronic alcoholics and poly-substance abusers compared with normal controls.

As demonstrated in the literature presented, many modifications and improvements have been reported for the VR subtest of the WMS-R. In summary, the following elements have added additional functionality to the VR subtest and are often used in clinical practice: a recognition trial or cued recall, matching and copy trials, the calculation of percentage of information retained, and supplemental scoring of intrusion errors.

1.9.5 Scoring

The scoring procedures for the original Visual Reproduction designs were criticised for their ambiguity (Heilbronner, 1992). The rules for scoring each design were not explicit, which resulted in considerable disagreement between raters and inconsistency evaluating the reproductions. The inter-rater reliability of the Visual Reproduction subtests from the Wechsler Memory Scale was formally examined by Crosson, Hughes, Roth and Monkowski (1984) who reported that correlations were all statistically significant, but some were only moderate in size. They also reported that large score discrepancies occurred, mostly because of differences of opinion regarding the required degree of accuracy. They concluded that further refinement of the scoring system for the Visual Reproduction subtest was warranted.

These previous revisions of the Visual Reproduction subtest in the WMS-R included scoring improvements, such as: a greater range of possible scores, more explicit scoring criteria, and more robust normative data. The scoring principles were further clarified by the inclusion of detailed drawing examples. The more refined scoring procedures, which were outlined in greater objective detail in the WMS-R, resulted in a reported improved inter-scorer agreement (Smith et al., 1992; Wechsler, 1987). Despite these improvements, there were still some fundamental limitations in the scoring of the WMS-R Visual Reproduction subtest, which subsequently limited its ability to accurately yield visual memory measures.

The standard Visual Reproduction scoring criteria, however, received several criticisms for remaining problems. The major limitations of the scoring procedure evident in clinical applications of WMS-R Visual Reproduction subtest were documented in the works of Loring (1989) and Loring and Papanicolaou (1987). They reported concern about the limited range of possible scores on three of the four designs, and noted that the limited range was likely to result in ceiling effects in young adults and floor effects in the elderly and the memory impaired, particularly on delayed recall. They identified that the scoring guidelines did not adequately cover the varied design range seen in clinical practice, and observed a difficulty in discrimination of relatively mild differences in performance. A further limitation of the scoring procedure was the absence of an explicit rationale for the number of total points allocated to each design; and as norms are only provided for the total score on the four designs, it was difficult to evaluate performance on a single specific design. An additional criticism was that the scoring of individual items was not entirely independent, resulting in consequential

errors where failure on one criterion can affect earning points on later criteria. Finally, the scoring criteria were seen as focusing significantly on attention to drawing detail, giving excessive weighting to factors other than memory. This creates problems when assessing memory where limited motor skill might lead to overshoots, imprecision or mild distortions, and these errors would be penalised in the scoring system and inappropriately treated as imperfections in memory.

1.9.6 Development of an alternate scoring system for the VR subtest of the WMS-R

In 2000, Dowling and Clark developed an alternate scoring system for the Visual Reproduction subtest of the WMS-R. The alternate scoring system was designed to increase its sensitivity in measuring non-verbal memory function, and address the limitations identified in the original WMS-R scoring system. The alternate scoring system aimed to reduce anomalies in scoring criteria, while simultaneously generating a similar grading of memory performance to improve the subtest's clinical utility and diagnostic value. The alternate scoring system aimed to reduce the emphasis on non-memory factors to enhance the quality and consistency of scoring with the alternate scoring system. It was anticipated that the Alternate Scoring System would generate a fairly similar overall grading of memory performance to that derived from the original WMS-R scoring criteria, whilst providing a better representation of non-verbal memory function.

Although a number of changes were made to the scoring system, there were no changes to the standard administration of the subtest. The alternate scoring system was developed on the following specific guidelines and principles (Clark, 2000):

1. Each design would be given equal weighting and allocated the same number of maximum points, e.g., twenty, to avoid subjective judgment about comparative difficulty and to permit comparison of performance on individual designs
2. Criteria would be scored entirely independently from each other
3. To minimise floor effects and to ensure that partial recall was rewarded, the criteria would start off at a low level, imperfect recall of the designs would still score points, and minor overshoots or imperfects would attract no penalty
4. Specific tolerances for angles, line lengths, minor gaps and curves would be consistent across designs
5. Criteria would be included for perfect design reproductions and for perfect reproductions not containing any additional elements

6. The criteria would address different aspects of the design rather than a few key elements and aim to cover the diversity of reproductions seen in clinical practice.

1.9.6.1 Reliability and correlational analysis of the alternate scoring system

An initial evaluation of the psychometric properties of this Alternate Scoring System was undertaken to test the hypotheses that the ASS would generate a similar grading of memory to the original scoring system and have sound inter-rater and intra-rater reliability that would be equal to, or better than, the original scoring system of the WMS-R (Dowling & Clark, 2000). Expectations were that the WMS-R criteria would substantially correlate with the revised criteria and that inter-rater reliability would be improved due to the reduction of non-memory factors impacting on the scoring.

Clark and Dowling (2000) obtained immediate and delayed recall protocols for the WMS-R Visual Reproduction subtest from 60 participants, aged 50-87 years. Thirty participants had evidence of neurological impairment documented by neurological examination or neuroimaging with the remaining participants forming a control group.

Intra-rater reliability for the alternate scoring system was examined by one author who scored the sample of 60 test protocols on both the original and alternate scoring systems on two separate occasions, approximately a fortnight apart. The results indicated strong intra-rater reliability with correlation coefficients of .99 for immediate and delayed recall on the alternate scoring system , a result better than that for the OSS (.96 and .97 respectively).

Inter-rater reliability for the alternate scoring system was examined by two authors both scoring thirty of the protocols drawn randomly from the sample. The inter-rater reliability was also uniformly very high with correlation coefficients for the immediate and delayed recall on the alternate scoring system of .95 and .99, respectively. This was compared with the correlation coefficients for the immediate and delayed recall on the OSS of .96 and .97, respectively.

Overall, the two scoring systems were found to share a high degree of common variance. In the control group, the correlation between the alternate and original scoring systems was .88 for immediate recall and .93 for delayed recall. In the clinical group, the correlation between the revised and original scoring systems was .92 for immediate

recall and .99 for delayed recall. The high correlation between the two scoring systems indicated a comparable grading of memory performance, with the alternate scoring system having certain advantages and fewer anomalies.

The literature presented has identified that the clinical utility of the WMS-R Visual Reproduction subtest is improved by the use of the Alternate Scoring System . The Clark and Dowling (2000) study showed that the Alternate Scoring System correlated highly with the original, indicating it had retained the integrity of the original WMS-R, whilst allowing more emphasis to be placed on non-verbal aspects of the test allowing for a better assessment of non-verbal memory function.

1.10 Rationale of the Current Study

In spite of its documented limitations, the Visual Reproduction subtest of the Wechsler scales is arguably one of the most widely used tests of non-verbal memory. The profusion of research and clinical data available on the VR subtest of Wechsler Memory Scale-Revised continues to makes it a popular measure amongst many clinicians in the neuropsychological assessment of memory.

Therefore, rather than develop an entirely new test of non-verbal memory, many authors have chosen instead to modify this existing test with the aim of producing a score that more accurately reflects non-verbal memory function. Indeed, several attempts have previously been reported specifically relating to modification of the VR subtest of the WMS-R (Dowling & Clark, 2000; Fastenau, 1996; Gass, 1995; Hanger et al., 1991; Haut et al., 1994; Haut et al., 1996; Wilhelm, 1996). Interestingly, only one of the published modifications or supplemental procedures designed to expand the clinical utility of the VR subtest of the WMS-R has included a revision of the standard scoring criteria (Dowling & Clark, 2000).

Overall, the preliminary evidence of the alternate scoring system developed by Clark (2000) appeared to address a number of the limitations of the original scoring system of the Visual Reproduction subtest of the Wechsler Memory Scale-Revised. Furthermore, the substantial literature review has indicated that the alternate scoring system for the VR subtest of the WMS-R will increase the clinical utility of the WMS-R when assessing visual memory.

However, although the research of Clark and Dowling (2000) details an improvement in scoring the WMS-R Visual Memory subtest, the correlational analysis reassures us that the test remains similar to the previous test, but does not indicate its comparability within the Australian population. It is a widely recognised fact that psychological tests assessing cognitive function require periodic revision of their content. Revisions of the original WMS have maintained its wide acceptance and also addressed the recognised importance of updating test standardisation (Flynn, 1998). However, to procure an accurate interpretation of test results, the requisition of reliable normative data has also been identified (Mitrushina et al., 2005; Spreen & Strauss, 1998). Furthermore, the importance of establishing local normative data to accurately assess memory functioning within specific populations has also been emphasised (Ivnik et al., 1992; Levin et. al., 1987; Mittenberg et al., 1992).

The Macquarie University Neuropsychological Normative Study (MUNNS) was conducted in the late 1990's with the aim of providing local normative data for commonly used neuropsychological assessment measures, including the WMS-R (Carstairs & Shores, 2000). The published normative data for an Australian population aged 18 to 34 years provided an advantage to utilising the WMS-R for this cohort in Australia. The MUNNS normative data were published to provide national standards against which the test performances of brain-injured Australian patients could be directly compared. Analysis of the MUNNS data suggested that US norms were not identical to Australian norms and that results needed to account for gender and demographic variables (Shores & Carstairs, 2000), hence the clinical relevance of the current project. In order for the validated alternate scoring system for the VR subtest of the WMS-R to be clinically useful, appropriate normative data is requisite. Provision of normative data is required to improve the clinical utility of the WMS-R in the Australian population.

1.11 Aims and Hypotheses

The first, and preliminary, aim of the study was to refine the wording of the original alternate scoring system in collaboration with the original authors to maximise intra-rater reliability and inter-rater reliability.

The second, and main, aim of the study was to produce normative data by applying the alternate scoring system to subject responses on the VR subtest of the WMS-R.

The hypotheses of this study are as follows:

1. Overall, the refined scoring criteria of the alternate scoring system (ASS) will generate a similar grading of memory, with a moderate-high positive correlation, with the original scoring system (OSS) of the Wechsler Memory Scale-Revised.
2. The ASS will be psychometrically sound, and have equal, or greater than, intra-rater reliability and inter-rater reliability, compared with the OSS.
3. The ASS will have demonstrate satisfactory criterion validity with the OSS.
4. The ASS will demonstrate satisfactory construct validity by discriminating between a clinical and control group.

METHOD

2.1 Study 1

2.1.1 The Macquarie University Neuropsychological Normative Study

The Macquarie University Neuropsychological Normative Study (MUNNS) was conducted in the late 1990's with the aim of providing local normative data for commonly used neuropsychological assessment measures, including the WMS-R (Carstairs & Shores, 2000). The author of the current study was not involved in any data collection as part of the MUNNS.

2.1.1.1 Participants

The MUNNS was conducted between January 1996 and March 1998, and involved 399 healthy adults who were recruited from Sydney, Australia. Participants were aged between 18 and 34 years, with a mean age of 25.64 years (SD = 4.97 years). There were 206 (51.6%) females and 193 (48.4%) males, overall participants had a mean of 12.93 years of education (SD = 2.05 years).

2.1.1.2 Sample design and recruitment of participants

The aim of the MUNNS was to provide normative data on commonly used neuropsychological measures used in the assessment and rehabilitation of brain-injured patients. Eighteen to 34 year-olds were chosen as the range, because they incorporated the normative reference group (20-34 year-olds) of the WMS-R, and included 18 and 19 year-olds as they are commonly involved in TBI and therefore a useful group to study.

The sampling procedure, designed in collaboration with the Australian Bureau of Statistics (ABS), was designed to be representative of 1991 census data for 18 to 34 year-olds residing in the Sydney metropolitan region. A stratified random sampling procedure of Sydney's regions ensured balanced participant selection for sex, age, non-English speaking background, socio-economic level, and educational background. Over 10,000 potential participants were contacted through random numbers derived from a computerised telephone book. The sampling and recruiting procedures of the study and an assessment of the representativeness of the MUNNS sample are described in detail elsewhere (Carstairs & Shores, 2000).

2.1.2 Procedure and measures

2.1.2.1 Background information questionnaire

Participants were asked over the phone to provide demographic and background information prior to testing to establish which individuals fulfilled the sample requirements. The questionnaire pertained to age, years of education, educational qualifications, occupation, language background, history of injury and loss of consciousness, and use of medicinal and recreational drugs. Any endorsement of a head injury resulting in unconsciousness, current therapeutic or recreational drug use that might impair task performance, inability to read or understand English, or a physical or intellectual disability that would prevent test completion, resulted in the participant's exclusion from the study.

2.1.2.2 Testing

Of the 1,270 respondents who fulfilled the sample requirements and who agreed to participate in the study and were contacted later to make an appointment for a three-hour testing session, only 399 completed the testing. Informed consent was obtained prior to the commencement of test sessions. The WAIS-R (Wechsler, 1981), the WMS-R (Wechsler, 1987), and eight to ten other neuropsychological tests were administered. Full details of the complete testing battery are outlined in Carstairs and Shores (2000). Personal transportation was provided for participants to and from the University where the testing was conducted, and registered psychologists administered all tests. Participants were paid $100 each for their contribution.

2.1.2.3 Visual Reproduction subtest

Administration of the Visual Reproduction subtest from the WMS-R was included in the test battery and was completed by all 399 subjects. Each participant's response on their scoring sheet was scored twice by independent assessors. The stimulus cards from the WMS-R can be found in Appendix A.

2.1.3 Obtaining MUNNS data set

For normative data to be representative and robust it requires, at least, several hundred subject responses. Given that the aim of the present study was to provide data for up to 400 subjects, it was considered beyond the scope of this particular study for the author to individually assess each participant. Instead, with permission, an electronic copy of the data set for the 399 participants of the original MUNNS study was obtained. The data set was in SPSS format, and contained all the participant demographic variables

and results from all of the tests administered, including the WAIS-R and WMS-R subtest and index scores.

2.1.4 Protocols

The author of the current study obtained permission from the researchers of the MUNNS project to gain direct access to the original protocols that the participants had completed. In 2002, over the course of two interstate trips to Macquarie University, Sydney, the author, firstly, sorted through the files for each participant and photocopied the summary score sheet of the WAIS-R. Secondly, the entire protocol booklet for the WMS-R was also photocopied for each participant. Finally, all of the Visual Reproduction response sheets, containing the hand drawings, for each participant were photocopied. After duplicating, the photocopies of the participants' drawings were individually screened to assess their quality. Where the quality of the photocopy was deemed to be unsatisfactory, annotations were marked on the photocopy and in several instances; a transparency was used to directly trace the participants' responses.

2.1.5 The Alternate Scoring System

2.1.5.1 Refinement of the wording

The wording of the scoring criteria in the alternate scoring system for the Visual Reproduction subtest of the WMS-R developed by Dowling and Clark (2000) underwent some initial revision. Refinements to scoring criteria wording were done in consultation with the original authors and were undertaken to reduce confusion, to assist with clarification and involved only minor adjustments to the wording, rather than a derivation from the spirit of the scoring system that was initially developed. These slight modifications were made to ensure reliable application of the alternate scoring criteria while preserving the original intention of the scoring rule.

2.1.5.2 Scoring

The refined scoring criteria of the alternate scoring system were applied to each of the 399 participants' responses. This was conducted entirely by the current author. Systematic scoring took place over twelve months, around late 2002 and early 2003. On average, up to 40 protocols were scored each week.

2.1.5.3 Scoring drift

When one scorer applies the same scoring criteria to multiple responses, the scorer may gradually, even unconsciously, begin to accept less (or sometimes demand more) than is

appropriate in awarding that particular score point. This process of 'scoring drift' can result in an inequitable situation where one particular subject's response could receive a different score depending on when the response was scored. To prevent 'drift' and maintain the consistency and accuracy of the scores, the author referred to protocols produced in the development of the alternate scoring system as examples of the various score points. These examples were used as 'anchors' because they assisted in fixing the acceptable range within a score point and prevented the scorer from 'drifting' higher or lower in expectations for awarding a score point.

2.1.5.4 Data entry

Once the original protocols and data set were obtained, further information was entered into an SPSS database by the current author. Although the Visual Reproduction subtest raw scores were already present in the data set obtained, these values were re-calculated and re-entered by the author from the protocols to ensure quality of the final dataset. In addition, the raw scores for each of the four designs, both immediate and delayed, which were not present in the original data set obtained, were also entered. Scores generated from applying the alternate scoring system to the Visual Reproduction subtest were also entered. Including: immediate recall and delayed recall, for each of the four designs, in addition to, total immediate and delayed recall scores.

2.1.5.5 Intra-rater reliability

In order to establish the reliability of scoring on both systems, intra-rater reliability was examined using the MUNNS study sample. During the original development of the alternate scoring criteria, intra-rater reliability was measured to be very high (rho = .94-.99; Clark, 2000). Due to the large number of protocols and the author's increased familiarity with scoring criteria, further data reliability checks were performed to curtail the process of scoring drift and ensure adequate intra-rater reliability of scoring. Several series of protocols were sampled and re-scored to check their reliability, with a total number of 40 protocols scored twice by the current author. Greater emphasis and attention were placed on re-scoring earlier protocols in the sample, because of the assumption that these were more likely to exhibit discrepancy over the passage of scoring time. Measures of consistency that were calculated to evaluate this form of reliability included correlational analysis and the percentage of agreement in the total score. A measure of internal consistency, Cronbach's Alpha, was also calculated. These measures were calculated between the scores collected on two separate occasions by the author.

2.1.5.6 Inter-rater reliability

In order to establish the reliability of scoring on both systems, inter-rater reliability was examined using the MUNNS study sample. During the original development of the alternate scoring criteria, inter-rater reliability was measured to be very high (rho = .87-.99; Clark, 2000). As minor modifications to the alternate scoring system protocol wording were made as part of this current study, extensive consultation with the alternate scoring system authors deemed that these changes would not be likely to significantly impact the previously obtained inter-rater reliability scores. However, several series of protocols were scored both by the current author and re-scored by a qualified neuropsychologist to check their reliability, with a total number of 40 protocols re-scored. Greater emphasis and attention were placed on re-scoring earlier protocols in the sample, because of the assumption that these were more likely to exhibit discrepancy over the passage of scoring time. Measures of consistency that were calculated to evaluate this form of reliability included correlational analysis and the percentage of agreement in the total score. A measure of internal consistency, Cronbach's Alpha, was also calculated. These measures were calculated between the scores obtained by the two separate raters.

2.1.6 Design

The experimental design was a single factor, within subjects design.

2.1.6.1 Variables

The dependent variable was the measured visual memory level of functioning of each participant. This was operationally defined as the visual reproduction test raw score. This dependent variable was continuous with an interval level of measurement. The independent variable was the scoring system, which was operationally defined by the type of scoring criteria applied to obtain a visual reproduction test raw score. The independent variable was a nominal variable with a discrete level of measurement, taking two levels: original scoring system and alternate scoring system.

2.2 Study 2

2.2.1 Participants

The study sample comprised 34 adults aged between 18 and 39 years. These participants were assigned to either a control group or an clinical group based on the absence or presence of a neurological condition.

2.2.1.1 Clinical group

The clinical group was comprised of 17 participants, all of whom were assessed as outpatients in a Melbourne neuropsychological private practice between April and November 2011. Each participant had documented evidence of cerebral dysfunction through neurological and/or neuroimaging examination. Clinical diagnoses were varied and diverse, and included: epilepsy, stroke, acquired brain injury, encephalitis, and aneurysm. Criteria for exclusion in the study included: the presence of significant auditory or visual impairment, severe aphasia, and assessment within the previous six months using any of the test materials used in the study. Participants were aged between 18 and 39 years, with a mean age of 27.76 years (SD = 7.18 years). There were 7 (41.2%) females and 10 (58.8%) males, overall participants had a mean of 12.29 years of education (SD = 1.61 years).

2.2.1.2 Control Group

Seventeen participants comprised the control group. They were chosen from the MUNNS as controls for the individuals in the clinical group. Participants were aged between 18 and 34 years, with a mean age of 26.88 years (SD = 5.93 years). There were 7 (41.2%) females and 10 (58.8%) males, overall participants had a mean of 11.88 years of education (SD = 1.80 years).

2.2.2 Materials

The materials used in this study included the Visual Reproduction subtest of the Wechsler Memory Scale-Revised (WMS-R),

2.2.1.2 Visual Reproduction subtest (Wechsler Memory Scale-Revised)

The Visual Reproduction subtest of the Wechsler Memory Scale-Revised (WMS-R: Wechsler, 1987) is a measure of visual or non-verbal memory. It was previously described in detail in Section 1.9.

The standard procedure for administering the Visual Reproduction subtest of the WMS-R, as outlined in the WMS-R manual, was used in this study. It consists of four cards with printed visual designs of increasing complexity, and which involved both immediate and delayed recall trials. The immediate recall trial, required participants to draw, from memory, of each of the four designs immediately following a 10 second presentation. The delayed recall trial, administered around half an hour later, required participants to, once again, draw from memory, of each of the previously presented

designs. The participant's recall for immediate and delayed trials were scored using both the original scoring system specified in the WMS-R manual and the alternate scoring system.

2.1.6 Design

The experimental design was a single factor, between subjects design.

2.1.6.1 Variables

The dependent variable was the measured visual memory level of functioning of each participant. This was operationally defined as the visual reproduction test raw score. This dependent variable was continuous with an interval level of measurement. The independent variable was the clinical/control group, which was operationally defined by the allocation to the clinical or control group based on the presence or absence of a neurological condition. The independent variable was a nominal variable with a discrete level of measurement, taking two levels: original scoring system and alternate scoring system.

RESULTS

3.1 Statistical analysis

Data were analysed using the Statistical Package for the Social Sciences (SPSS) for Windows (Versions: 10.0.5, 1999 & 20.0, 2011).

3.1.1 Data Analysis

All variables were inspected for skewness and kurtosis. Skewness and kurtosis values revealed that many variables were negatively skewed or were flat in distribution. Conformity to parametric assumptions was formally examined with Kolmogorov-Smirnov normality tests. In many cases the critical value of the Kolmogorov-Smirnov test exceeded .05, indicating departure from normality. Thus, many of the variables did not conform to the assumptions underlying parametric statistics and therefore, a conservative approach was taken to data analysis. When conducting the correlation analyses Spearman's Rho was calculated and when analyses were conducted to compare two groups Mann-Whitney U statistics were used.

3.2 Test development

3.2.1 Refinement of the Alternate Scoring System

The refinement of the alternate scoring system (ASS) was based on the principles outlined in the design. Consistent with the original development of the alternate scoring system, the general scoring rule that stated that all items were to be scored independently of each other was retained. In total, fourteen of the eighty items underwent revision. The unrevised wording of the alternate scoring system criteria and corresponding changes are presented in Appendix B. The resultant wording of the alternate scoring system, comprising of the 20 scoring criteria for each of the four designs are presented in Appendices C-F.

3.3 Reliability

The term reliability is used to describe the consistency of a score on a particular test across testing and/or scoring situations. The reliability of the alternated scoring system (ASS) and original scoring system (OSS) was examined in terms of:

1. the consistency of scoring as applied by one rater on two occasions (intra-rater),
2. the consistency in scoring between two raters (inter-rater), and
3. the internal consistency of the Alternate Scoring System.

3.3.1 Intra-rater reliability

Intra-rater reliability refers to the consistency of a single rater in scoring the same test on two separate occasions. In order to examine intra-rater reliability, the same rater (i.e., the author) rated a sample of 40 test protocols from the MUNNS study sample on both the original scoring system and alternate scoring system on two occasions. The variables examined included the score on each of the four designs and the total score. These variables were scored for both immediate and delayed recall, on both the alternate scoring system and the original scoring system across the two scoring occasions.

In order to examine the consistency of scoring across the two separate occasions, a number of measures were calculated. These measures included a correlational analysis to examine the degree of association between the scores obtained on the two occasions and the percentage of agreement in the score for each design across the two scoring occasions. Both of these analyses will now be described in detail.

3.3.1.1 Correlational analysis

Traditionally, correlation studies have been used to explore intra-rater reliability. By examining the relationship between scores collected across two separate occasions, high correlation coefficients are seen as reflecting a high degree of consistency in scoring. Thus, a correlation analysis was conducted on the set of 40 protocols scored on two occasions by the one rater. Spearman's correlations were calculated as the primary measure of intra-rater reliability due to significant skewness and/or kurtosis in many of the variables. The correlation coefficients for immediate and delayed recall are presented in Table 3.1.

Spearman's correlation coefficients for the alternate scoring system were uniformly high, ranging from *rho* = .95 for Design One, Immediate Recall; to *rho* = .99 for Designs Four and Two, Delayed Recall. Coefficients for the original scoring system in this sample were generally lower, ranging from *rho* = .91 for Design Two, Immediate Recall; to *rho* = .98 for Design One, Delayed Recall. However, the differences between two scoring systems were not significant and all coefficients were very high.

The very strong correlations reported above suggested that there was a high level of consistency between the scores obtained on the first occasion and the scores obtained on a separate occasion two weeks later. Thus, it appeared that an individual rater could score both scoring systems with a high degree of reliability.

Table 3.1

Correlations between scoring of Immediate and Delayed Recall on Two Occasions

	Alternate Scoring System		Original Scoring System	
	rho	Sig.	rho	Sig.
Immediate Recall				
Design 1	.95	.0005	.94	.0005
Design 2	.98	.0005	.91	.0005
Design 3	.98	.0005	.96	.0005
Design 4	.99	.0005	.93	.0005
Total	.99	.0005	.97	.0005
Delayed Recall				
Design 1	.97	.0005	.98	.0005
Design 2	.99	.0005	.97	.0005
Design 3	.98	.0005	.97	.0005
Design 4	.96	.0005	.97	.0005
Total	.99	.0005	.98	.0005

N = 40

3.3.1.2 Scoring Agreement

Correlation analysis can provide an estimate of the degree of association between two scores obtained on two separate occasions. However, it does not provide direct information about the actual agreement in the total score obtained for each individual design across the two separate scoring occasions (although it can imply a high level of agreement). That is, it does not provide direct information as to whether, for example, a score of 18 on one design corresponds to a score of 18 on the same design on the second scoring occasion. Thus, for each of the four designs, a 'percentage of score agreement' method was used to examine intra-rater reliability across the two scoring occasions for each protocol. An error was defined as a non-identical item score between the first and second scorings. A retest of the sample of forty test protocols resulted in a detailed comparison of 6400 items.

As some error in measurement was also to be expected due to random factors, a cut-off level for acceptable agreement in the total score was established. Agreement in the total score within one point was set as an acceptable level of agreement (for e.g. a score of 18 on the first occasion and a score of 19 on the second occasion for the same design would illustrate acceptable agreement). As the total score for each design on the

alternate scoring system was 20 points, a one-point variation was 5% of the total score on each design. This highly conservative criterion was equal to the traditional 5% cut-off often used in psychological research. The results of the intra-rater reliability samples are presented in Table 3.2

Table 3.2

Intra-rater Reliability (number of non-identical item scores)

	N	Design 1	Design 2	Design 3	Design 4	TOTAL	% errors[a]
Sample 1	10	5	21	17	6	49	3.1
Sample 2	10	4	23	19	10	56	3.5
Sample 3	10	7	7	7	6	27	1.7
Sample 4	10	4	5	8	2	19	1.2
TOTAL	40	20	56	51	24	151	2.4%

[a] Percentage of non-identical item scores as a total of all items

Comparison results indicated a small proportion of 'scoring drift' evident in the earlier two samples (3.1% and 3.5%), which was slightly higher than compared with the last two samples (1.7% and 1.2%). Overall, the total percentage of non-identical item scores was 2.4%, which indicated high intra-rater reliability and was deemed to be well below the 5% criterion, and therefore satisfactory for the purposes of the current study. Data for the intra-rater reliability analysis can be found in Appendix G.

3.3.2 Inter-rater reliability

In tests where scoring requires some judgement, it is important to examine the extent to which reliability might be affected by variation in this judgement between raters. In order to examine the consistency in scoring between different raters, two independent raters scored 40 of the test protocols drawn from the sample of 399 reported above.

One of these raters was the author and the other was a qualified neuropsychologist who was not involved in the development or revision of the alternate scoring system. Although the raters were aware of the purpose of their scoring, they were blind to the other raters scores. No identifying information was provided about the participants who had provided the protocols for scoring.

Each rater scored immediate and delayed recall for each of the four designs on both the original scoring system and the alternate scoring system. The variables included in the

analysis were the score for each of the four designs and the total score. In order to examine the consistency of scoring across the two separate raters, a number of measures were calculated. These measures included a correlational analysis to examine the degree of association between the scores obtained by the two raters and the percentage of agreement in the score for each design across the two raters. Both of these analyses will now be described in detail.

3.3.2.1 Correlational analysis

Correlation studies have been used to explore inter-rater reliability. By examining the relationship between scores collected between the two separate scorers, high correlation coefficients are seen as reflecting a high degree of consistency in scoring. Thus, a correlation analysis was conducted on the set of 40 protocols scored by the two raters. Spearman's correlations were calculated as the primary measure of inter-rater reliability due to significant skewness and/or kurtosis in many of the variables. The correlation coefficients for immediate and delayed recall are presented in Table 3.3.

Table 3.3

Correlations between scoring of Immediate and Delayed Recall by Two Raters

	Alternate Scoring System		Original Scoring System	
	rho	Sig.	rho	Sig.
Immediate Recall				
Design 1	.96	.0005	.83	.0005
Design 2	.92	.0005	.89	.0005
Design 3	.98	.0005	.82	.0005
Design 4	.99	.0005	.93	.0005
Total	.98	.0005	.94	.0005
Delayed Recall				
Design 1	.95	.0005	.81	.0005
Design 2	.95	.0005	.85	.0005
Design 3	.95	.0005	.91	.0005
Design 4	.97	.0005	.87	.0005
Total	.98	.0005	.96	.0005

N = 40

Spearman's correlation coefficients for the alternate scoring system were uniformly high, ranging from *rho* = .92 for Design Two, Immediate Recall; to *rho* = .99 for Design

Four, Immediate Recall. Coefficients for the original scoring system in this sample were generally lower, ranging from *rho* = .81 for Design One, Delayed Recall; to *rho* = .93 for Design Four, Immediate Recall. However, the differences between two scoring systems were not significant and all coefficients were high.

The very strong correlations reported above suggested that there was a high level of consistency between the scores obtained from the first rater and the scores obtained from the second rater. Thus, it appeared that two individual raters could score either scoring system with a high degree of reliability.

3.3.2.2 Scoring Agreement

Correlation analysis can provide an estimate of the degree of association between two scores obtained on two separate occasions. However, it does not provide direct information about the actual agreement in the total score obtained for each individual design across the two separate raters (although it can imply a high level of agreement). That is, it does not provide direct information as to whether, for example, a score of 18 on a design from one rater corresponds to a score of 18 on the same design from the second rater. Thus, for each of the four designs, a 'percentage of score agreement' method was used to examine inter-rater reliability across the two raters for each protocol. An error was defined as a non-identical item score between the two raters. A retest of the sample of forty test protocols resulted in a detailed comparison of 6400 items.

As some error in measurement was also to be expected due to random factors, a cut-off level for acceptable agreement in the total score was established. Agreement in the total score within one point was set as an acceptable level of agreement (for e.g. a score of 18 from the first rater and a score of 19 from the second rater for the same design would illustrate acceptable agreement). As the total score for each design on the alternate scoring system was 20 points, a one-point variation was 5% of the total score on each design. This highly conservative criterion was equal to the traditional 5% cut-off often used in psychological research. The results of the inter-rater reliability samples are presented in Table 3.4

Table 3.4

Inter-rater Reliability (number of non-identical item scores)

	N	Design 1	Design 2	Design 3	Design 4	TOTAL	% errors[a]
Sample 1	10	15	26	23	15	79	4.9
Sample 2	10	12	28	14	9	63	3.9
Sample 3	10	18	9	16	8	51	3.2
Sample 4	10	8	16	7	5	36	2.3
TOTAL	40	53	79	60	37	229	3.6%

[a] Percentage of non-identical item scores as a total of all items

Similar to the results of the intra-rater reliability, comparison results indicated a small proportion of 'scoring drift' evident in the earlier two samples (4.9% and 3.9%), which was slightly higher than compared with the last two samples (3.2% and 2.3%). Overall, the total percentage of non-identical item scores was 3.6% which indicated high inter-rater reliability and was deemed to be below the 5% criterion, and therefore satisfactory for the purposes of the current study. As compared to the total percentage of non-identical item scores for intra-rater reliability, there would appear to be marginally lower agreement between two raters as compared to one rater, although the level of agreement was still very high. Data for the inter-rater reliability analysis can be found in Appendix H.

3.3.2 Internal Consistency

The potential reliability of the alternate scoring system was also measured by evaluating the internal consistency of the scoring items for each of the four designs. The internal consistency of the Alternate Scoring System for immediate and delayed recall was evaluated by computing Cronbach's alpha. The internal consistency of the four designs was computed using the data from MUNNS study sample (N = 399). The results are presented in Table 3.5 for each of the Visual Reproduction designs across the Alternate Scoring System.

Table 3.5

Internal consistency of the Alternate Scoring System for immediate and delayed recall

	Cronbach's Alpha	
	Immediate	Delayed
Design 1	.88	.97
Design 2	.83	.91
Design 3	.67	.68
Design 4	.86	.94
Total	.91	.93

N = 399

Table 3.3 shows the internal consistency reliability coefficient for the Immediate Recall scores of Designs One, Two and Four was high, but only moderate reliability for Design Three was evident. The internal consistency of the Total Immediate Recall score was also high which suggested good internal reliability of the Alternate Scoring System in this study. Similarly, the delayed recall internal consistency coefficients for Designs One, Two and Four were high, but only moderate reliability for Design Three was evident. The internal consistency of the Total Delayed Recall score was high and reflected favourably on the reliability of the Alternate Scoring System.

3.4 Normative data

3.4.1 Sample characteristics

The MUNNS normative sample encompassed 399 participants with a mean age of 25.64 years (SD = 4.97 years, with 206 (51.6%) females and 193 (48.4%) males and a mean of 12.93 years of education (SD = 2.05 years). Final sample characteristics are presented in Tables 3.6 and 3.7.

Table 3.6

Sample Characteristics: Age Range

	18-21 years	22-25 years	26-29 years	30-34 years	Total
	112	89	89	109	399
Total	28.1%	22.3%	22.3%	27.3%	(100%)

N = 399

Table 3.7

Sample Characteristics: Gender and Years of Education

	8-12 years	13-19 years	Total
Male	89	104	193 (48.4%)
Female	93	113	206 (51.6%)
Total	182 (45.6%)	217 (54.4%)	399 (100%)

N = 399

3.4.2 Sample intelligence quotient distribution

Figure 3.1 presents the frequency distribution of the WAIS-R full scale intelligence quotient (FSIQ) scores for the sample. The statistical properties of the FSIQ scores for the sample were computed, including the mean, standard deviation, variance, kurtosis, and skewness. These results are presented in Table 3.8.

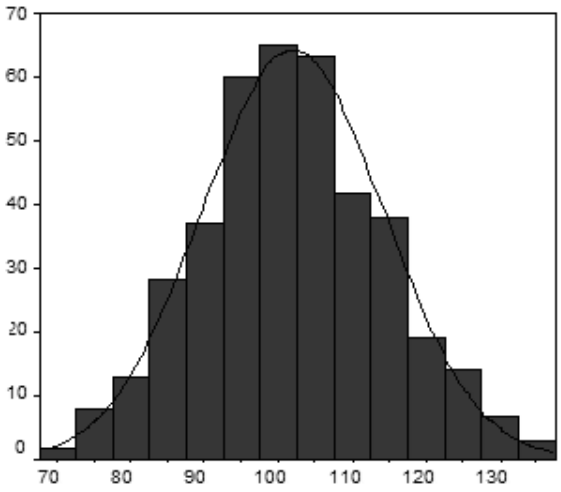


Figure 3.1 Frequency Distribution of the MUNNS sample WAIS-R Full Scale IQ Scores

Table 3.8

WAIS-R Full Scale IQ Scores: Descriptive Statistics

	M	SD	Variance	Kurtosis	Skewness
Full Scale IQ	101.93	12.36	152.72	-.17	.126

N = 399

The data presented in Figure 3.1 and Table 3.8 show that the FSIQ scores for the sample were normally distributed. Data of the descriptive statistics for the original scoring system and alternate scoring system can be found in Appendix I.

3.4.3 Scoring systems distributions

3.4.3.1 Distribution of scores for the original scoring system

Figures 3.2 and 3.3 present the frequency distribution of the original scoring system for immediate and delayed recall. The statistical properties of the data upon which the original scoring system was applied were computed, including the mean, standard deviation, variance, kurtosis, and skewness. These results are presented in Table 3.9.

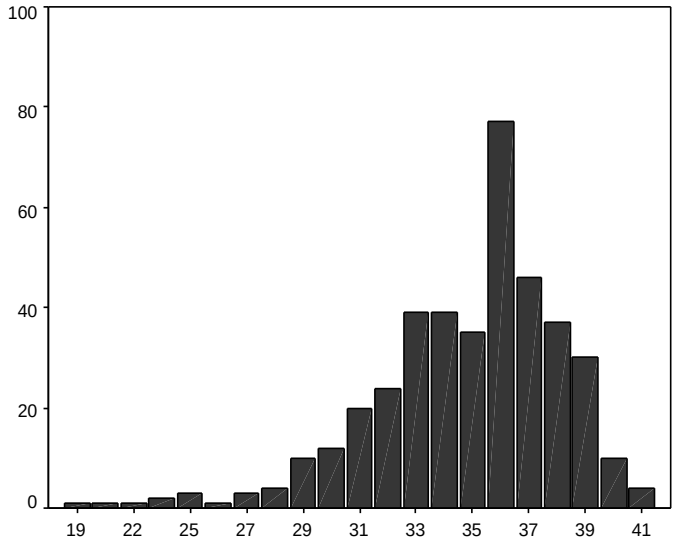


Figure 3.2 Distribution of Original Scoring System - Immediate Recall Total Scores

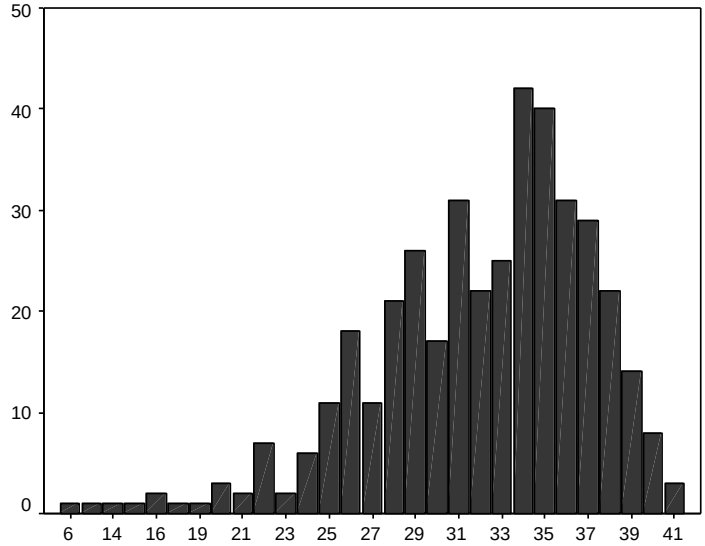


Figure 3.3 Distribution of Original Scoring System - Delayed Recall Total Scores

Table 3.9

Original Scoring System: Descriptive Statistics

	M	SD	Variance	Kurtosis	Skewness
Immediate Recall					
Design 1	5.80	0.86	0.74	1.42	-0.82
Design 2	5.48	1.29	1.66	-0.72	-0.40
Design 3	7.36	1.31	1.72	2.50	-1.34
Design 4	16.19	1.82	3.33	2.93	-1.55
Total	34.82	3.35	11.25	2.19	-1.10
Delayed Recall					
Design 1	4.90	2.15	4.63	1.11	-1.58
Design 2	4.58	2.05	4.21	0.27	-0.95
Design 3	7.07	1.61	2.60	4.21	-1.73
Design 4	15.49	2.66	7.06	8.08	-2.34
Total	32.05	5.16	26.65	2.02	-1.07

N = 399

Figures 3.2 and 3.3 and Table 3.9 show that both the immediate and delayed recall total scores for the original scoring system were not normally distributed. Inspection of the skewness and kurtosis revealed that both the immediate and delayed recall total score distributions were significantly negatively skewed with a flatter than a normal distribution. Indicating that fewer participants than predicted scored very poorly on both immediate and delayed total scores and that more participants than predicted scored in the regions in between the mean and the extremes.

3.4.3.2 Distribution of scores for the alternate scoring system

Figures 3.4 and 3.5 present the frequency distribution of the alternate scoring systems for immediate and delayed recall. The statistical properties of the alternate scoring system were computed, including the mean, standard deviation, variance, kurtosis, and skewness. These results are presented in Table 3.10.

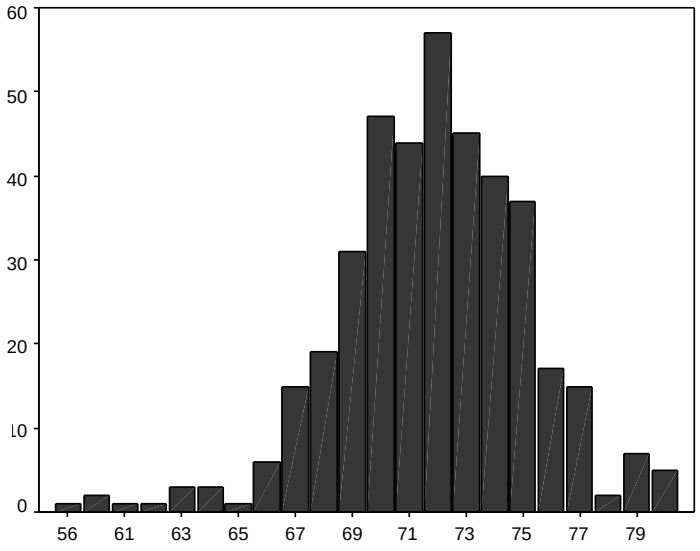


Figure 3.4 Distribution of Alternate Scoring System - Immediate Recall Total Scores

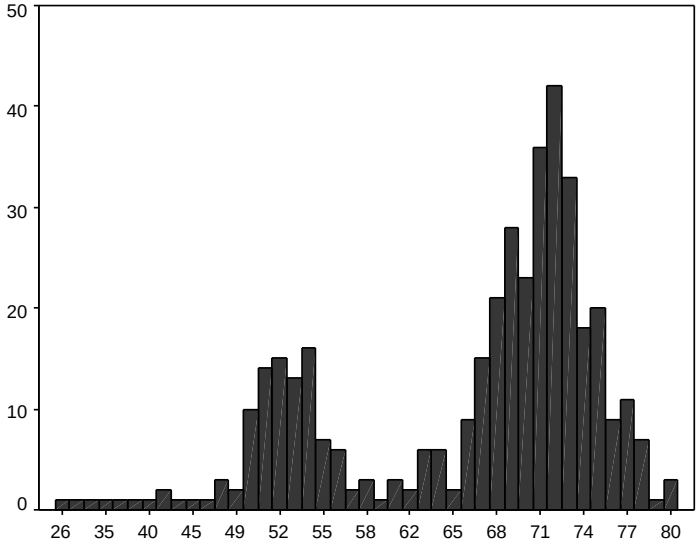


Figure 3.5 Distribution of Alternate Scoring System - Delayed Recall Total Scores

Table 3.10

Alternate Scoring System: Descriptive Statistics

	M	SD	Variance	Kurtosis	Skewness
Immediate Recall					
Design 1	18.15	0.91	0.82	0.22	-0.03
Design 2	17.99	1.42	2.00	0.34	-0.60
Design 3	17.12	1.88	3.55	2.46	-1.11
Design 4	18.13	1.47	2.17	3.44	-1.27
Total	71.39	3.79	14.35	1.23	-0.65
Delayed Recall					
Design 1	15.43	6.37	40.52	2.03	-1.97
Design 2	15.71	5.68	32.28	3.46	-2.21
Design 3	16.63	2.79	7.79	15.06	-3.15
Design 4	17.61	2.35	5.51	18.80	-3.43
Total	65.39	9.74	94.80	0.81	-1.09

N = 399

Figures 3.4 and 3.5 and Table 3.10 show that both the immediate and delayed recall total scores for the alternate scoring system were not normally distributed. Inspection of the skewness revealed that both the immediate and delayed recall total score distributions were significantly negatively skewed and that the delayed recall total scores had a bi-modal distribution. Kurtosis measures also indicated a significant departure from a normal distribution, particularly for the delayed recall total scores. As with the original scoring system, fewer participants than predicted scored very poorly on both immediate and delayed total scores.

3.4.3.3 Scaled scores for the Alternate Scoring System

Scaled scores for the alternate scoring system were generated using percentiles. The scaled scores were based on a mean score of 10 (50th percentile) and a standard deviation of 3. These results are presented in Table 3.11. Data of the percentile ranks for the alternate scoring system can be found in Appendix J.

Table 3.11

Alternate Scoring System-Scaled Scores

Scaled Score	Immediate Recall Total Raw Score	Delayed Recall Total Raw Score	Scaled Score
19	80	80	19
18	79	79	18
17	.	.	17
16	78	78	16
15	77	77	15
14	76	75-76	14
13	75	73-74	13
12	74	72	12
11	73	71	11
10	71-72	67.-.70	10
9	69-70	58.-.66	9
8	68	53.-.57	8
7	67	51.-.52	7
6	65-66	49.-.50	6
5	63-64	47.-.48	5
4	61-62	38.-.45	4
3	59-60	31.-.37	3
2	57-58	27.-.30	2
1	57>	27>	1

3.5 Validity

Validity refers to the ability of a test to adequately assess the hypothetical construct it was designed to measure in different populations. This study examined the criterion and construct validity of the alternate scoring system for the Visual Reproduction subtest

3.5.1 Normality of the two scoring system distributions

Conformity to parametric assumptions was formally examined for both scoring systems with Kolmogorov-Smirnov normality tests, and in all cases the critical value of .05, indicated a significant departure from a normal distribution. These results are presented in Table 3.12. Data of the tests of normality for the OSS and ASS can be found in Appendix K.

Table 3.12

Kolmogorov-Smirnov [a] Normality of the Two Scoring System Distributions

	Original Scoring	Alternate Scoring	df	Sig.
Immediate Recall				
Design 1	0.29	0.25	399	.0005
Design 2	0.18	0.17	399	.0005
Design 3	0.24	0.16	399	.0005
Design 4	0.23	0.19	399	.0005
Total	0.15	0.10	399	.0005
Delayed Recall				
Design 1	0.33	0.41	399	.0005
Design 2	0.20	0.31	399	.0005
Design 3	0.23	0.20	399	.0005
Design 4	0.21	0.24	399	.0005
Total	0.12	0.20	399	.0005

[a] Lilliefors Significance Correction

N = 399

3.5.2 Correlations within two scoring systems

Tables 3.13 and 3.14 present the correlation coefficients between the subtests for both scoring systems. The correlations between the subtests comprising both the immediate and delayed recall for the original scoring system, although mostly significant, were very weak. The correlations between the subtests comprising both the immediate and delayed recall for the alternate scoring system were stronger than the original scoring

system (with the exception of Design 4, for both immediate and delayed recall). However, as with the original scoring system, the correlations between the subtests for the alternate scoring system, although mostly significant, were weak.

Table 3.13

Spearman's Rho correlations between the Original Scoring System Subtests

	Immediate Recall			Delayed Recall		
	Design 1	Design 2	Design 3	Design 1	Design 2	Design 3
Immediate						
Design 2	.12*					
Design 3	.15**	.19**				
Design 4	.07	.15**	.27**			
Delayed						
Design 2				.08		
Design 3				.21**	.18**	
Design 4				.11*	.22**	.32**

* Correlation is significant at the .05 level (2-tailed).

** Correlation is significant at the .01 level (2-tailed).

N = 399

Table 3.14

Spearman's Rho correlations between the Alternate Scoring System Subtests

	Immediate Recall			Delayed Recall		
	Design 1	Design 2	Design 3	Design 1	Design 2	Design 3
Immediate						
Design 2	.23**					
Design 3	.22**	.26**				
Design 4	.31**	.16**	.22**			
Delayed						
Design 2				.12*		
Design 3				.28**	.28**	
Design 4				.20**	.18**	.28**

* Correlation is significant at the .05 level (2-tailed).

** Correlation is significant at the .01 level (2-tailed).

N = 399

3.5.3 Criterion Validity

Criterion related validity is based on a tests' correlation with other tasks that measure similar processes. The criterion validity of the alternate scoring system, as compared with the original scoring system, was examined by through a correlational assessment of the similarity of grading between the two scoring systems. Figures 3.6 and 3.7 present scatter plots of the immediate and delayed recall total scores for both scoring systems. The scatter plots indicated reasonably linear relationships, therefore Spearman Rho correlation coefficients for the immediate and delayed recall subtests and total scores between both scoring systems were calculated. These results are presented in Table 3.15.

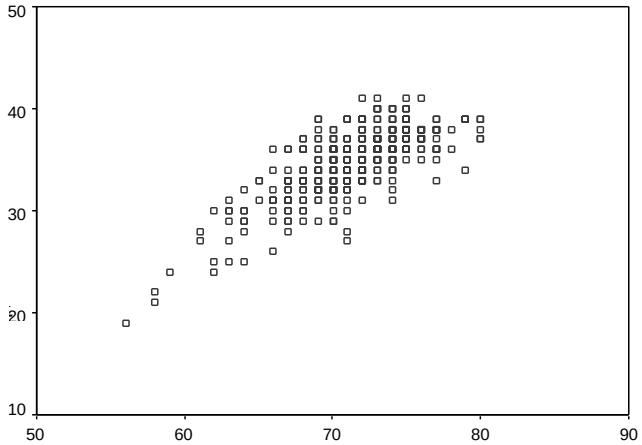


Figure 3.6 Scatter plot of Immediate Recall Total Scores for both Scoring Systems

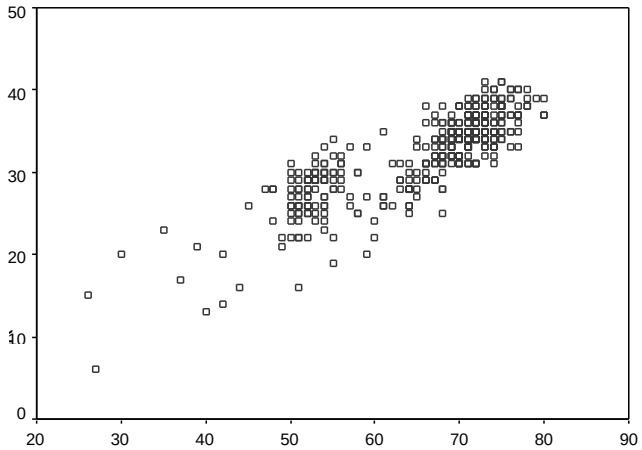


Figure 3.7 Scatter plot of Delayed Recall Total Scores for both Scoring Systems

Table 3.15

Correlations between the Alternate and Original Scoring Systems

	rho	Sig.
Immediate Recall		
Design 1	0.507	.0005
Design 2	0.386	.0005
Design 3	0.571	.0005
Design 4	0.525	.0005
Total	0.696	.0005
Delayed Recall		
Design 1	0.687	.0005
Design 2	0.548	.0005
Design 3	0.560	.0005
Design 4	0.619	.0005
Total	0.836	.0005

N = 399

The correlations for the subtests comprising both the immediate and delayed recall and total scores between both scoring systems were all positive and significant. There were moderately strong to strong sized correlations between the immediate (r = .70) and delayed recall (r = .84) total scores. Of the subtests comprising the immediate and delayed recall total scores, there were weak (r = .39) to moderately strong correlations (r = .69). Data of the nonparametric correlations can be found in Appendix L.

3.5.4 Construct Validity

This study also examined the construct validity of the alternate scoring system of the Visual Reproduction subtest. Construct validity refers to the extent that a test measures the theoretical construct of interest. It is often inferred by looking at group studies and seeing if a task is able to discriminate between clinically meaningful groups. In this study, the ability of both the alternate and original scoring systems to discriminate between a group of persons with known neurological conditions (the clinical group) and a group of persons with no evidence of lesions (the control group) was examined in order to provide information about construct validity. Given that validity was examined by looking at a clinical and a control group, the demographic characteristics of these groups were compared.

3.5.4.1 Demographic Characteristics of the Study Sample

Two groups of 17 participants, one group with documented cerebral dysfunction and a group with no evidence of cerebral dysfunction, were compared to see if they were suitably matched groups. The construct validity study sample comprised 34 adults aged between 18 and 39 years. These participants were assigned to either a control group or an clinical group based on the absence or presence of a neurological condition. Participants in the clinical group were aged between 18 and 39 years, with a mean age of 27.76 years (SD = 7.18 years). There were 7 (41.2%) females and 10 (58.8%) males, overall participants had a mean of 12.29 years of education (SD = 1.61 years). Participants in the control group were aged between 18 and 34 years, with a mean age of 26.88 years (SD = 5.93 years). There were 7 (41.2%) females and 10 (58.8%) males, overall participants had a mean of 11.88 years of education (SD = 1.80 years). The results of t-test comparisons between the dependent variable, group membership (control or clinical; male or female) and the independent variables (age, years of education, or gender) are presented in Table 3.16.

Table 3.16

T-test Comparisons Between Groups for Age, Years of Education and Gender

Group	N	Variable	*t*	df	*p*
Clinical x Control	34	Age	.695	32	.490
		Years of Education	-.175	32	.862
		Gender	-.510	32	.612
Males x Females	34	Age	-1.925	32	.059
		Years of Education	.538	32	.593
Males	20	Age	1.470	18	.154
Clinical x Control		Years of Education	-.686	18	.499
Females	14	Age	-.253	121	.802
Clinical x Control		Years of Education	.334	12	.740

There were no significant differences between the control and clinical groups in terms of age, years of education or gender composition. There were also no significant differences in age or years of education between males and females, between males in the clinical and control groups, or between females in the clinical and control groups. However, the difference in age between the males and females in this study did approach significance, with a trend for the female group to be slightly older.

3.5.4.2 Comparison between Clinical and Control Groups

Construct validity was examined by looking at the subtest's ability to discriminate between groups. In this way, construct validity was expected in the form of successful discrimination between two groups, the clinical group and the control group.

Mann-Whitney U tests were calculated to examine whether there were any significant differences between the control group and clinical group on their memory for designs as scored according to the alternate scoring system and original scoring systems. The results of the Mann-Whitney U comparison between the clinical and control groups are presented in Table 3.17. Data of Mean ranks and sum of ranks can be found in Appendix M. Significance was set at the .05 level.

With the exception of three comparisons: Designs One and Two, Immediate Recall, on the Original Scoring System; and Design One, Immediate Recall on the Alternate Scoring System, the scores obtained on all designs were significantly different between the control and clinical groups. In all significant cases, the control group had a higher mean rank score than the clinical group, indicating that the clinical group performed more poorly than the control group in their memory for designs.

Table 3.17

Mann-Whitney U Comparisons Between the Clinical and Control Groups

	z	df	*p*
Alternate Scoring System			
Immediate Recall			
Design 1	-1.735	32	.083
Design 2	-2.245	32	.025
Design 3	-3.774	32	.000
Design 4	-3.900	32	.000
Total Score	-3.780	32	.000
Delayed Recall			
Design 1	-2.256	32	.024
Design 2	-4.076	32	.000
Design 3	-2.847	32	.004
Design 4	-2.512	32	.012
Total Score	-3.838	32	.000
Original Scoring System			
Immediate Recall			
Design 1	-1.145	32	.252
Design 2	-1.408	32	.159
Design 3	-3 .648	32	.000
Design 4	-3.698	32	.000
Total Score	-3.924	32	.000
Delayed Recall			
Design 1	-2.327	32	.020
Design 2	-3.740	32	.000
Design 3	-2.541	32	.011
Design 4	-2.380	32	.017
Total Score	-3.346	32	.000

N = 34

DISCUSSION

Comprehensive neuropsychological examination requires an accurate assessment of memory function. Knowledge about the integrity of memory systems can contribute to important decisions regarding differential diagnoses, rehabilitation strategies, competency status, and surgery options. The Wechsler Memory Scales have dominated the clinical assessment of adult memory functioning over the past sixty years. Since its conception, the Visual Reproduction (VR) subtest has been a core element that has been retained in subsequent revisions and represents a widely accepted measure of non-verbal memory function. The broad application of the VR subtest in clinical practice has led to significant and prolific research investigation findings, including: the identification of numerous limitations (Heilbronner, 1992; Loring, 1989; Loring & Papanicolaou, 1987; Troester et al., 1993) and several documented attempts at varied modifications designed to improve and enhance the clinical utility of the subtest (Fastenau, 1996; Gass, 1995; Hanger et al., 1991; Haut et al. 1996; Haut et al. 1994; Wilhelm, 1996). Of the many modifications and improvements suggested in the literature, including: a recognition trial or cued recall; matching and copy trials; the calculation of percentage of information retained; and supplemental scoring of intrusion errors, several were found to add further functionality to the VR subtest but none proposed an entirely revised scoring structure.

The present study extended the work of Dowling and Clark (2000) who initially developed an alternate scoring system (ASS) for the VR subtest of the WMS-R. Their aim was primarily to address the multifarious problems and significant limitations identified in published research with the original scoring system (OSS) for the VR subtest of the WMS-R. A limitation of particular concern, for example, was that within the OSS the items were not applied independently. This resulted in the unintended, and undesired, effect of the contribution of a single key item possibly inflating the overall score or that portions of recall may not have been rewarded or acknowledged at all. Consequently, the ASS was designed so that each item could be scored independently. In this way the ASS made fewer qualitative assumptions about the way information was retained and focused more specifically on grading the degree of actual memory recall. The ASS was also designed to allow for a greater examination of drawing quality, in terms of considering of what was present rather than strictly what was absent. The ASS was not intended to be a radical departure from the OSS, rather it was conceptualised as a scoring system that contained fewer anomalies but still reflected a similar grading of memory.

4.1 Refinement of the ASS

The first, and preliminary, aim of the study was to refine the wording of the original alternate scoring system, in collaboration with the original authors, to maximise intra-rater reliability and inter-rater reliability. In order for the alternate scoring system to be clinically useful in the assessment of memory it must have good reliability. Furthermore, to be considered as a genuine alternative to the original scoring system, the alternate scoring system must be at least equal, in reliability to the Visual Reproduction subtest of the Wechsler Memory Scale-Revised.

For the current project, fourteen of the eighty original ASS scoring items underwent minor wording revision. Refinements to the wording of the ASS items were based upon principles outlined in the study's design and were consistent with the original development of the alternate scoring system. The modifications were made to improve intra-rater reliability and additionally, also inter-rater reliability. Wechsler (1987) himself improved the wording of the scoring criteria of the VR subtest similarly when he revised the WMS with an aim of minimising scoring complexity whilst maximising scoring accuracy.

4.1.1 Intra-rater reliability

The reliability of the alternate scoring system (ASS) and original scoring system (OSS) was examined in terms of intra-rater reliability. In the present study intra-rater reliability was examined using two measures: correlational analysis to examine the degree of association between the scores obtained on the two occasions, and the percentage of score agreement in the score for each design across the two scoring occasions.

Correlational analysis indicated consistently high intra-rater reliability for both scoring systems indicating that an individual rater could score both scoring systems with a high degree of reliability. In general, the coefficients for the ASS indicated greater intra-reliability when compared with the OSS; however, the differences between two scoring systems were not significant and all coefficients were very high.

A 'percentage of score agreement' method was used to examine intra-rater reliability across the two scoring occasions for each protocol. An error was defined as a non-identical item score between the first and second scorings. It was expected, and indeed was found, that there was a larger percentage of 'scoring drift' evident in the earlier two samples as compared to the latter two samples of the study. The total overall percentage

of non-identical items was 2.4%, indicating a reasonably high intra-rater reliability, which was found to be satisfactory for the purpose of this study given its aim of producing normative data. Overall, the results of the intra-rater reliability analysis also provided support for the hypothesis that the ASS would have equal, or greater, intra-rater reliability than the OSS.

The low percentage of errors and high intra-rater reliability reported for the modified ASS in the present study was comparable to the high intra-rater reliability initially reported for the VR subtest of the WMS-R by Wechsler (1987) in a normative sample population and later by Woloszyn, Grob-Murphy, Wetzel, and Fisher (1993) in a clinical population. These findings of the present study, that the interpretation and application of the ASS was stable over time, contribute to the enhanced clinical relevance and improved utility of the VR subtest across of broad subsection of patient groups.

4.1.2 Inter-rater reliability

The reliability of the alternate scoring system (ASS) and original scoring system (OSS) was also examined in terms of inter-rater reliability. In the present study inter-rater reliability was examined using two measures: correlational analysis to examine the degree of association between the scores on each design obtained by two separate raters using the ASS, and the percentage of agreement in the score for each design between the two raters.

Correlational analysis indicated consistently high inter-rater reliability for both scoring systems indicating that separate raters could score both scoring systems with a high degree of reliability. In general, the coefficients for the ASS indicated greater inter-reliability when compared with the OSS; however, the differences between two scoring systems were not significant and all coefficients were high. It should be noted however, that the ASS contained almost double the number of items of the OSS (80 items vs. 41 items), therefore a greater potential variability existed on the ASS. If the increased number of items on the ASS is taken into consideration, the similar degree of high reliability indicates that the ASS may have slightly better inter-reliability. Nevertheless, it is important to consider that the original scoring system can have high inter-rater reliability when applied rigorously.

During the original development of the alternate scoring criteria, inter-rater reliability was reported to be very high (*rho* = .95-.99; Dowling & Clark, 2000). Although minor modifications to the ASS protocol wording were made as part of this current study, extensive consultation with the ASS authors ensured that these changes would not significantly impact the previously obtained high inter-rater reliability scores.

A 'percentage of score agreement' method was used to examine inter-rater reliability across the scores from the two raters. An error was defined as a non-identical item score between the first and second rater. Similar to the results of the intra-rater reliability, it was expected, and indeed was found, that there was a larger percentage of 'scoring drift' evident in the earlier two samples as compared to the latter two samples of the study. The total overall percentage of non-identical items was 3.6%, indicating reasonably high inter-rater reliability, which was found to be satisfactory for the purpose of this study given its aim of producing normative data. Overall, the results of the inter-rater reliability analysis provided support for the hypothesis that the ASS would have equal, or greater, inter-rater reliability than the OSS.

The inter-rater reliability of the original scoring system (OSS) has previously been investigated by its author, Wechsler (1987), and others (Woloszyn et al., 1993). The results from the present study were similar to that Wechsler's (1987) normal sample and the clinical group of Woloszyn et al. (1993). Wechsler (1987) reported inter-rater reliability coefficients of .97 in the original standardisation sample of the WMS-R and Woloszyn et al. (1993) reported reliability coefficients of .98 for immediate recall and .98 for delayed recall in a clinical population. These results are generally consistent with the findings of the current study.

High levels of inter-rater reliability on the Visual Reproduction subtest have been reported in several studies. However, larges differences between the raw scores obtained across scorers on individual designs have also been reported. Wechsler (1987), in publishing original inter-rater reliability data for the subtest, reported that the raw score for each design varied by four or less points across raters. He did report that large differences were infrequent and that the average total raw score difference across raters was only 1.5 points.

4.1.3 Internal consistency

The reliability of the alternate scoring system (ASS) and original scoring system (OSS) was also examined in terms of internal consistency. Results of the internal consistency of the ASS were high and reflected favourably on the reliability of the Alternate Scoring System. Estimates were .91 for the total immediate recall and .93 for total delayed recall. Overall, the results of the internal consistency analysis provided further support for the hypothesis that the ASS would be psychometrically sound, with equal, or greater, intra-rater and inter-rater reliability than the OSS.

Previously reported internal consistency estimates for the VR subtest of the WMS-R ranged from .46 to .71 for immediate recall and .38 to .59 for delayed recall (Wechsler, 1987; Williams et al, 1998). It is suggested that the increased the reliability of the ASS could be due, in part, to the greater number of items for each of the four designs.

4.2 Production of normative data

The importance of establishing local normative data to accurately assess memory functioning within specific populations has previously been emphasised (Ivnik et al., 1992; Levin et. al., 1987; Mittenberg et al., 1992). Furthermore, specific concerns about the uptake of the Wechsler Memory Scales in Australia due to the absence of local normative data have previously been documented (Holdnack, Lissner, Bowden & McCarthy, 2004). It was therefore considered that in order for the validated ASS for the VR subtest of the WMS-R to be clinically useful in memory assessment, apposite normative data was required.

The present study aimed to produce normative data by applying the ASS to subject responses on the VR subtest of the WMS-R collected through the MUNNS. The sample characteristics in this study were 399 participants with a mean age of 25.64 years. The young adults sampled were considered to represent a satisfactory balance of gender and covered a reasonable range of years of age and years of education. An examination of the sample's intelligence quotient indicated a normal distribution, further supporting the acceptable representativeness of the sample.

As reported previously, the immediate and delayed recall total scores from the OSS were not normally distributed. An inspection of both the skewness and kurtosis showed that immediate and delayed recall total score distributions were significantly negatively skewed with a flatter than normal distribution, indicating that fewer participants in the

study than predicted scored very poorly on both immediate and delayed total scores. Furthermore, more participants than predicted scored in regions in between the mean and extremes. The study also demonstrated that similar to the OSS, both the immediate and delayed recall total scores for the ASS were also abnormally distributed. Inspection of the skewness revealed that both the immediate and delayed recall total score distributions for the ASS were also significantly negatively skewed and that the delayed recall total scores had a bi-modal distribution. Kurtosis measures also indicated a significant departure from a normal distribution, particularly for the delayed recall total scores. As with the OSS, fewer participants than predicted scored very poorly on both immediate and delayed total scores. The conformity to parametric assumptions was formally examined and indicated a significant departure from a normal distribution.

The abnormal distribution of both scoring systems and the findings that fewer participants than predicted scored very poorly on both immediate and delayed total scores was thought to be accounted for by the deliberate exclusion of participants with: a history of head injury; current drug use, deemed likely to impair task performance; illiteracy; an inability to understand English; and physical or intellectual disability that would prevent test completion. It is standard practice to screen out those individuals with deficits when forming a normative sample. A normative sample, by definition, is comprised of those individuals whose function is assumed to be 'normal'. Inclusion of individuals with specific deficits can significantly lower the overall mean and negatively skew the normative data. Collection of data for particular clinical groups with specific deficits is a separate process that occurs once a 'normal' population have been examined, so that the baseline of normal function can be used as a reference point for comparison. Nonetheless, it was encouraging that both scoring systems demonstrated a similarly skewed distribution.

Scaled scores for the alternate scoring system were generated using percentiles and were chosen as the format for the normative data due to their general clinical utility. By reporting the scaled scores for the normative sample the second, and main, aim of the study: to produce normative data by applying the alternate scoring system to subject responses on the VR subtest of the WMS-R, was achieved.

4.3 Further validation of ASS in comparison to OSS

Correlational analysis for the subtests contained within both scoring systems was conducted. Although significant correlations were anticipated between the designs

comprising both the immediate and delayed recall for the OSS, the correlations were mostly significant but very weak. By comparison, the correlations between the subtests comprising both the immediate and delayed recall for the ASS were slightly stronger than the OSS (with the exception of Design 4, for both immediate and delayed recall). However, as with the OSS, the correlations between the subtests for the ASS, although mostly significant, were also weak. Nonetheless, the finding of improved correlations can be seen as indicative of a more cohesive and internally consistent set of scoring procedures. This finding supported the initial aim of Clark and Dowling (2000) to create a scoring system that retained the original purpose and integrity of the test, whilst allowing for some alterations in scoring emphasis. This finding also further supported the preliminary aim of the present study to make only slight modifications to ensure reliable application of the ASS whilst preserving the original intention of the scoring rules.

4.3.1 Criterion Validity

Through a comparison of the ASS and the OSS, the hypothesis was examined that the ASS would generate a similar grading of memory to the OSS but that it would have equal or greater criterion validity, reflecting an analogous grading of memory. The results further demonstrated that reasonably linear relationships were indicated and accordingly, correlation coefficients for the immediate and delayed recall subtests and total scores between both scoring systems were calculated. The correlations for the subtests comprising both the immediate and delayed recall and total scores between both scoring systems were all positive and significant. These findings supported the hypothesis that the alternate scoring system would generate a similar grading of memory and would have equal or greater criterion validity to the original scoring system, and provided further evidence of the alternate scoring system's test validity and indicated the successful retention of the original purpose and integrity of the test while allowing some change in scoring emphasis. Furthermore, there were moderately strong to strong sized correlations between the immediate and delayed recall total scores which further supported the hypothesis that the refined scoring criteria of the alternate scoring system (ASS) would generate a similar grading of memory, with a moderate-high positive correlation, with the original scoring system (OSS) of the Wechsler Memory Scale-Revised.

Of the subtests comprising the immediate and delayed recall total scores, there were weak to moderately strong correlations. Interestingly, these results were significantly

lower than the correlations for immediate recall and delayed recall reported by Clark (2000). The lower correlation between the two scoring systems could be perceived to be directly related to the refinements to the wording of fourteen items in the ASS resulting in a less comparable grading of memory performance; however, the process of modification was conducted in careful collaboration with the original authors to ensure that the integrity of each item of the ASS was retained. Alternatively, these low correlations were thought to be more likely due to a limited dispersion resulting from a truncated range. A spuriously low correlation coefficient can be the product of a truncated range, which is a condition where the range of values is restricted. Given the generally high level of functioning of the participants in the MUNNS study and, therefore, the greater capacity of the subjects to verbally encode the design, this may possibly reflect the ability of the participants, as a group, to easily make semantic associations between elements of the design. A sample group with a broader range of IQ scores and a much larger sample size may have demonstrated performances on the designs that more closely resembled a normal distribution.

In the present study the age range of 18 to 34 years was studied, which does not include the ages where the greatest normal changes in memory function occur and where the incidence of nervous system abnormalities increases (Lezak, et al, 2004). It may be possible that given a more generous sample of age ranges that higher correlations may have been obtained. Although the MUNNS sample did not include a breadth of ages to allow for exploration of age related changes that would be expected in an aging population, the young adults studied were a group who are expected to have the best performance on memory tests, resulting in a suitable cohort upon which to base normal performances.

4.3.2 Construct Validity

The ability of the ASS and OSS to discriminate between clinically meaningful groups was examined. In this study, the groups defined as: persons with known neurological conditions (the clinical group) and persons with no evidence of lesions (the control group) were examined in order to provide information about construct validity. Demographic characteristics of these groups were compared, which indicated no significant differences between the control and clinical groups in terms of age, years of education or gender composition

Results indicated a very high degree of successful discrimination. The scores obtained on all designs were significantly different between the control and clinical groups, with the exception of three comparisons: Designs One and Two, Immediate Recall, on the Original Scoring System; and Design One, Immediate Recall on the Alternate Scoring System. In all significant cases, the control group had a higher mean rank score than the clinical group, indicating that the clinical group performed more poorly than the control group in their memory for designs. The finding that the ASS was able to successfully discriminate between the clinical and control group on all designs for delayed recall; and on all designs, except one, for immediate recall; which was a better performance than the OSS, providing support for the hypothesis that the ASS would demonstrate satisfactory construct validity by discriminating between a clinical and control group. The stronger construct validity results for delayed recall designs, compared with immediate recall designs, has also been reported in previous studies (Larrabee & Curtiss, 1995; Lee, Loring, & Thompson, 1989). This was thought, in part, to be attributable to the stronger degree of similarity in scoring when any actual delayed recall for individual designs is absent (i.e., a design is completely 'forgotten' between the immediate and delayed recall trials. This occurrence can increase the likelihood of obtaining satisfactory construct validity results. Therefore, the results for the immediate recall designs are best seen to represent the true validity of the scoring systems. Regardless, the ASS did provide a more accurate and robust tool for discrimination between clinically significant groups.

4.4 Problems and Limitations of the Study

Methodological aspects of the present study that require consideration include the composition of the normative sample group. The sample selection of the MUNNS was deliberately biased to eliminate participants who were likely to exhibit significant memory dysfunction and was necessarily limited in magnitude and scope. These factors were considered acceptable with the aim of producing robust normative data; however, the spuriously low correlation coefficients obtained between the two scoring systems in the present study were considered to be directly attributable to the limited sample size and span. However, although the sample contained a biased selection of participants, it is standard practice to screen out those individuals with deficits when forming a normative sample. A normative sample, by definition, is comprised of those individuals whose function is assumed to be 'normal'. Inclusion of individuals with specific deficits can significantly lower the overall mean and negatively skew the normative data. Collection of data for particular clinical groups with specific deficits is a separate

process that occurs only after a 'normal' population have been examined, so that the baseline of normal function can be used as a reference point for comparison. Therefore, the external validity of generalising the results of the present study to a broad range of people and situations should be approached with caution. Notwithstanding, however, the resultant normative data of the present study is considered to be sufficient to be used immediately in Australian clinical practice.

Furthermore, the interpretation of performances of the clinical group on the Visual Reproduction subtest may also be misleading. The restricted sample characteristics of the clinical group may also limit the generalizability of performances as this was not a stratified and randomised study. Such that, participants in this study were recruited via convenience sampling which can be potentially misleading and biased towards education, ethnicity and socio-economic status (Holdnack et al, 2004). However, it should be noted that the sample sizes of the Wechsler Memory Scale-Revised were small with around 50 to 55 participants in each of the six age bands, that were used to generate the normative data. The statistical power from a sample of 34 participants was considered sufficient to draw meaningful conclusions from the data, even though the sample size in this study was relatively small (Cohen, 1988; Tabachnick & Fidell, 1996).

4.5 Future Directions

Future research directions building upon the work of the present study are recommended to include collation of normative data for a wider age range. As outlined previously, the VR subtest has been reported to sensitive to differentiating typical forgetting seen in an aging population compared with abnormal forgetting associated with a cortical dementia (Cullum, Butters, Troester & Salmon, 1990). Normative data specifically addressing an elderly population would likely be particularly relevant clinically, as this cohort are known to suffer the greatest incidence of age-related memory decline and there is an increasing need for assessment tools to accurately differentiate between the expected memory loss associated with age-related decline compared with insidious pathology. Furthermore, the psychometric properties of the ASS should be investigated using a sample with a wider IQ range, particularly in the lower range of intellectual functioning.

Additionally, future avenues of research could explore the potential benefits of separate norms for each individual design, allowing for detailed examination of trends in recall for individual designs which are currently obscured in the final score. Normative data

for separate designs may also be used to identify which design, if any, is an accurate and more reliable measure of non-verbal memory. Fatigue can represent a significant risk to cognitive endurance in lengthy assessments in vulnerable clinical populations, particularly the elderly. As such, the administration of individual designs may provide a more clinically parsimonious assessment in addressing the referral question in certain cases (Lacritz & Cullum, 2003). This may facilitate the development of a shortened test version; an outcome that would likely be considered advantageous in a time-poor clinical setting.

It would be advantageous to determine information about the capacity of the ASS to make finer distinctions about recall quality. Although the results of the present study indicated that the two scoring systems produce a similar grading of memory performance, the results of this study cannot firmly determine that the ASS necessarily produces a better grading of memory recall. Future avenues of investigation could possibly include experienced clinicians, naïve to the purpose of the study, grading the quality of memory recall based on clinical judgement and then comparing this to scores obtained on the OSS and the ASS.

As verbal processes have been identified as being used when people process non-verbal information, the development of a non-verbal index study has the potential to be able to identify participants with primarily non-verbal memory impairment. An index of scoring items that are performed more poorly by individuals with non-verbal memory deficits could be developed to potentially discriminate between clinical populations and quantify the severity of non-verbal memory impairment.

4.6 Conclusion

The importance of accurately assessing memory functioning is an essential and integral part in the framework of any comprehensive neuropsychological examination. The present study has reported on the modification and validation of an alternate scoring system (ASS), originally proposed by Clark and Dowling (2000), for a test of non-verbal memory: the Visual Reproduction subtest of the Wechsler Memory Scale-Revised. The literature review of the present study, in addition to supporting the development of the modified alternate scoring system, demonstrated the importance of Australian normative data collection for commonly used neuropsychological measures, particularly for use in the assessment and rehabilitation of brain-injured adult patients.

The process by which Australian normative data was generated for the modified alternate scoring system was also outlined in the present study.

When assessing the validity of the ASS as an improved alternative to the original scoring system (OSS) of the WMS-R, this research found that the ASS generated a similar grading of memory to the OSS; however, the modifications resulted in a more cohesive and internally consistent set of scoring procedures. This improvement in the instrument's reliability contributes to the enhanced clinical relevance and utility of the VR subtest across of broad subsection of potential adult patient groups.

The ASS, together with the normative data produced through the present study, provides another tool for the assessment of visual memory functioning in an Australian population and specifically responds to the reported concerns surrounding the use of the Wechsler Scales in Australia due to the absence of local normative data. It was noted in this study that there was only a weak correlation between the scoring systems; however, this was thought to be a result of the restricted range which did not include the ages where the greatest normal changes in memory function occur. The sample used within the present study was an age group expected to have the best performance on memory tests, consequently given the select group of this study it is not possible to comment on age related changes in typical memory functioning.

This study has accomplished its aim of production of Australian normative data for an alternate scoring system for the Visual Reproduction subtest of the WMS-R. The impact of this normative data is substantial as the Wechsler Memory Scales represent one of the most commonly used assessment tools of adult memory in everyday neuropsychological practice. Furthermore, accurate assessment and subsequent knowledge about the integrity of memory functioning can contribute to important decisions regarding differential diagnoses, competency, rehabilitation strategies and surgery options. In conclusion, the modification and validation of this improved scoring method for the VR subtest combined with the collation of local normative data form a significant contribution to knowledge, and extension of the evidence base for the practice of, clinical neuropsychology in Australia.

Recommendations for future research include a specific focus toward the development of normative data for an elderly population. This research has also emphasised the potential benefits of separate normative data for each design, allowing for examination

of trends in recall for individual designs which are currently obscured in the final score, and potential for a shortened test version.

REFERENCES

Adeyemo, S.A. (2002). A review of the role of the hippocampus in memory. *Psychology and Education: An Interdisciplinary Journal,* 39(2), 46-63.

Aggleton, J.P., & Saunders, R.C. (1997). The relationships between temporal lobe and diencephalic structures implicated in anterograde amnesia. *Memory,* 5(1-2), 49-72.

Alessio, A., Pereira, F.R., Sercheli, M.S., Rondina, J.M., Ozelo, H.B., Bilevicius, …Cendes, F. (2011). Brain plasticity for verbal and visual memories in patients with mesial temporal lobe epilepsy and hippocampal sclerosis: An fMRI study. *Human Brain Mapping, (Online version of record published before inclusion in an issue).*

Anderson, S. W., Damasio, H., & Tranel, D. (1993). Impaired memory for temporal order (recency) following ventromedial frontal lobe damage. *Society for Neuroscience Abstracts,* 19, 791.

Appollonio, I. M., Grafman, J., Schwartz, V., Massaquoi, S., & Hallett, M. (1993). Memory in patients with cerebellar degeneration. *Neurology*, 43(88), 1536-1544.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Baddeley, A. D. (1994). Working Memory: The interface between memory and cognition. In D. L. Schacter & E. Tulving (Eds.), *Memory Systems* (pp. 352 – 368). Cambridge, MA: MIT Press.

Baddeley, A. D. (1995). The psychology of memory. In A. D. Baddeley, B. A. Wilson, & F. N. Watts (Eds.), *Handbook of Memory Disorders.* England: John Wiley & Sons Ltd.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Bakker, A., Kirwan, C.B., Miller, M., & Stark, C.E.L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science,* 319(5870), 1640-1642.

Barr, W.B., Chelune, G.J., Hermann, B.P., Loring, D.W., Perrine, K., Strauss, E., Trenerry, M.R., & Westerveld, M. (1997). The use of figural reproduction tests as measures of nonverbal memory in epilepsy surgery candidates. *Journal of the International Neuropsychological Society*, 3, 435-443.

Baxter, M. G., & Murray, E. A. (2001). Impairments in visual discrimination learning and recognition memory produced by neurotoxic lesions of rhinal cortex in rhesus monkeys. *European Journal of Neuroscience,* 13, 1228 – 1238.

Bear, M.F., Connors, B.W., & Paradiso, M.A. (1996). *Neuroscience: exploring the brain.* Baltimore: Williams and Wilkins.

Benedict, R.H.B., & Groninger, L. (1995). Preliminary standardisation of a new visuospatial memory test with six alternate forms. *The Clinical Neuropsychologist,* 9(1), 11-16.

Benton, A. L. (1992). *The Revised Visual Retention Test* (5th ed.). New York: Psychological Corporation.

Benton, A.L. (1974). *Revised Visual Retention Test: Clinical and experimental applications* (4th ed.). New York: The Psychological Corporation.

Bigler, E.D., Johnson, S.C., Anderson, C.V., Blatter, D.D., Gale, S.D., Russo, A.A., Ryser, D.K., McNamara, S.E., Bailey, B.J., Hopkins, R.O., & Abildskov, T.J. (1996). Traumatic brain injury and memory: the role of hippocampal atrophy. *Neuropsychology,* 10, 333-342.

Bliss, T. V. & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, 232, 331-356.

Bloom, B.L. (1959). Comparison of the alternate Wechsler Memory Scale forms. *Journal of Clinical Psychology*, 15(1), 72-74.

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology,* 24(3), 383-405.

Bornstein, R. A. (1982). Effects of unilateral lesions on the Wechsler Memory Scale. *Journal of Clinical Psychology,* 38(2), 389-392.

Bornstein, R. A., & Chelune, G. J. (1989). Factor structure of the Wechsler Memory Scale-Revised in relation to age and educational level. *Archives of Clinical Neuropsychology,* 4, 15 – 24.

Bornstein, R.A, &. Chelune, G.J. (1988). Factor structure of the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist,* 2, 107-115.

Bowden, S. C., Ritter, A. J., Carstairs, J. R., Shores, A. E., Pead, J., Greeley, J. D., Whelan, G., Long, C. M., & Clifford, C. C. (2001). Factorial invariance for combined *Wechsler* Adult Intelligence Scale – Revised and Wechsler Memory Scale-Revised scores in a sample of clients with alcohol dependency. *The Clinical Neuropsychologist,* 15(1), 69 – 80.

Brandt, J., Shpritz, B., Munro, C.A., Marsh, L., & Rosenblatt, A. (2005). Differential impairment of spatial location memory in Huntington's disease. *Journal of Neurology, Neurosurgery, and Psychiatry,* 76, 1516-1519.

Broadbent, N. J., Clark, R. E., Zola S., & Squire, L. R. (2002). The medial temporal lobe and memory. In L. R. Squire & D. L. Schacter (Eds.), *Neuropsychology of Memory* (pp. 3 – 23). New York: Guilford Press.

Buffalo, E.A., Reber, P.J., & Squire, L.R. (1998). The human perirhinal cortex and recognition memory. *Hippocampus,* 8, 330-339.

Burwell, R.D., Witter, M.P., & Amaral, D.G. (1995). Perirhinal and postrhinal cortices of the rat: a review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus,* 5, 390-408.

Busch, R. M., Booth, J. E., McBride, A., Vanderploeg, R. D., Curtiss, G., & Duchnick, J.J. (2005). Role of executive functioning in verbal and visual memory. *Neuropsychology,* 19, 171 – 180.

Butters, N. & Stuss, D. T. (1989). Diencephalic amnesia. In F. Boiler & J. Grafman (Eds.), *Handbook of neuropsychology.* Vol 3, (pp. 107-149). New York: Elsevier Science Publishers.

Butters, N. (1984). The clinical aspects of memory disorders: contributions from experimental studies of amnesia and dementia. *Journal of Clinical Neuropsychology,* 6, 17-36.

Butters, N., Delis, D. C., & Lucas, J. A. (1995). Clinical assessment of memory disorders in amnesia and dementia. *Annual Review of Psychology,* 46, 493 – 524.

Butters, N., Salmon, D. P., Cullum, C. M., Cairns, P., Troester, A. I., Jacobs, D., Moss, M., & Cermak, L. S. (1988). Differentiation of amnesic and demented patients with the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist,* 2, 133-148.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: Empirical review of 275 PET and fMRI studies. *The Journal of Cognitive Neuroscience,* 12, 1-47.

Cabeza, R., Ciaramelli, E., Olson, I.R., & Moscovitch, M. (2008). The parietal *cortex and episodic memory: an attentional account. Nature* Reviews: Neuroscience, 9(8), 613-625.

Cahill, L., Uncapher, M., Kilpatrick, L., Alkire, M.T., & Turner, J. (2004) Sex-related hemispheric lateralization of amygdala function in emotionally influenced memory: an fMRI investigation. *Learning & Memory*, 11, 261- 266.

Caplan, B., & Caffery, D. (1996). Visual form discrimination as a multiple-choice visual memory test: illustrative data. *The Clinical Neuropsychologist,* 10(2), 152-158.

Cardoso-Leite, P., & Gorea, A. (2010). On the Perceptual/Motor Dissociation: A Review of Concepts, Theory, Experimental Paradigms and Data Interpretations. *Seeing and Perceiving,* 23(2), 89-151.

Carstairs, J. R., & Shores, E. A. (2000). The Macquarie University Neuropsychological Normative Study (MUNNS): Rationale and methodology. *Australian Psychologist,* 35, 36-40.

Chelune, G. J., Naugle, R. I., Luders, H., & Awad, I. (1991). Prediction of cognitive change as a function of preoperative ability status among temporal lobe patients seen at 6-month follow-up. *Neurology,* 41, 399-40

Chelune, G.J., & Bornstein, R.A. (1988). Wechsler Memory Scale-Revised patterns among patients with unilateral brain lesions. *The Clinical Neuropsychologist,* 2, 121-132.

Clark, R.E., & Squire, L.R. (2010). An animal model of recognition memory and medial temporal lobe amnesia: history and current issues. *Neuropsychologia,* 48(8), 2234-2244.

Clark, T. (2000). Development of an alternate scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised. *(Unpublished Doctoral Thesis).*

Clark, T., & Dowling, P. (2000). A revised scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised. Part 2: psychometric properties, relationship to original scoring system, and development of a non-verbal memory index. *(Unpublished paper).*

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and pattern-analysing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210, 207-210.

Cohen, N.J., & Eichenbaum, H. (1993). *Memory, amnesia and the hippocampal system.* Cambridge, Massachusetts: The Massachusetts Institute of Technology Press.

Corkin, S., Amaral, D. G., Gonzalez, R. G., Johnson, K. A., & Hyman, B. T. (1997). H.M.'s medial temporal lobe lesion: Findings from magnetic resonance imaging. *Journal of Neuroscience,* 17, 3964 – 3979.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* 11, 671 – 684.

Crosson, B., Hughes, C.W., Roth, D.L., & Monkowski, P.G. (1984). Review of Russell's 1975 norms for the Logical Memory and Visual Reproduction subtests of the Wechsler Memory Scale. *Journal of Consulting and Clinical Psychology*, 52(4), 635-641.

Cullum, C.M., Butters, N., Troester, A.I. and Salmon, D.P. (1990). Normal aging and forgetting rates on the Wechsler Memory Scale-Revised. *Archives of Clinical Neuropsychology*, 5(1), 23-30.

Damasio, R. C, Graff-Radford, P., Eslinger, P. J., Damasio, H. & Kassell, N. (1985). Amnesia following basal forebrain lesions. *Archives of Neurology,* 42, 263-271.

Davis, M. (2006). Neural systems involved in fear and anxiety measured with fear-potentiated startle. *American Psychologist,* 61, 741-756.

de Haan, M., Mishkin, M., Baldeweg, T., & Vargha-Khadem, F. (2006). Human memory development and its dysfunction after early hippocampal injury. *Trends in Neurosciences,* 29, 374-381.

Delis, D.C. (1989). Neuropsychological assessment of learning and memory. In Boller, F., & Grafman, J. (Eds.), *Handbook of Neuropsychology* (Vol 3, pp. 3-33). Amsterdam: Elsevier Science Publishers.

Desgranges, B., Baron, J.C., & Eustache, F. (1998). The functional neuroanatomy of episodic memory: the role of the frontal lobes, the hippocampal formation, and other areas. *Neuroimage,* 8(2), 198-213.

Desmond, J.E., & Fiez, J.A. (1998). Neuroimaging studies of the cerebellum: language, learning and memory. *Trends in Cognitive Sciences,* 2(9), 355–362.

Deweer, B, Pillon, B, Pochon, J. B., & Dubois, B. (2001). Is the H. M. story only a 'remote memory'? Some facts about hippocampus and memory in humans. *Behavioural Brain Research,* 127, 209 – 224.

Deweer, B., Lehericy, S., Pillon, B., Baulac, M., Chiras, J., Marsault, C., …Dubois, B. (1995). Memory disorders in probable Alzheimer's disease: the role of hippocampal atrophy as shown with MRI. *Journal of Neurology, Neurosurgery and Psychiatry,* 58(5), 590-597.

Dowling, P. and Clark, T. (2000). A revised scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale-Revised. Part 1: rationale, development and advantages. *(Unpublished paper).*

Dusoir, N., Kapur, D.P., Byrnes, S., McKinstry, R.D., & Hoare, R.D. (1990). The role of diencephalic pathology in human memory disorder. Evidence from a penetrating paranasal brain injury. *Brain,* 113(6), 1695-1706.

Echávarri. C., Aalten, P., Uylings, H.B., Jacobs, H.I., Visser, P.J., Gronenschild, …Burgmans, S. (2011). Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease. *Brain Structure and Function,* 215(3-4), 265-271.

Eichenbaum, H. (2002). *The cognitive neuroscience of memory: an introduction.* New York: Oxford University Press.

Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron,* 44, 109-120.

Elwood, R.W. (1991). Factor structure of the Wechsler Memory Scale-Revised (WMS-R) in a clinical sample: A methodological reappraisal. *The Clinical Neuropsychologist,* 5, 329-337.

Emilien, G., Durlach, C., Antoniadis, E., Van der Linden, M., & Maloteaux, J-M. (2004). *Memory: Neuropsychological, imaging and psychopharmacological perspectives.* New York: Psychology Press.

Erikson, R.C., & Scott, M.L. (1977). Clinical memory testing: a review. *Psychological Bulletin*, 84, 1130-1149.

Eslinger, P. J. (2002). *Neuropsychological interventions: Clinical research and practice.* USA: The Guilford Press.

Fahle, M., & Daum, I. (1997). Visual learning and memory as functions of age. *Neuropsychologia,* 35, 1583 – 1589.

Fastenau, P.S. (1996). An elaborated administration of the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist,* 10(4): 425-434.

Figueiredo, P., Santana, I., Teixeira, J., Cunha, C., Machado, E., Sales, F., …Castelo-Branco, M. (2008). Adaptive visual memory reorganization in right medial temporal lobe epilepsy. *Epilepsia,* 49(8), 1395-1408.

Fischer, J.S. (1988). Using the Wechsler Memory Scale-Revised to detect and characterize memory deficits in multiple sclerosis. *The Clinical Neuropsychologist,* 2(2), 149-172.

Fletcher, P.C., & Henson, R.N.A. (2001). Frontal lobes and human memory: insights from functional neuroimaging. *Brain,* 124(5), 849-881.

Flynn, J.R. (1998). WAIS-III and WISC-III gains in United States from 1972 to 1995: how to compensate for obsolete norms. *Perceptual and Motor Skills,* 86(3), 1231-1239.

Frankland, P.W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews: Neuroscience,* 6, 119-130.

Frisk, V., & Milner, B. (1990). The role of the left hippocampal region in the acquisition and retention of story content. *Neuropsychologia,* 28, 349 – 359.

Fuster, J.M (2008). *The Prefrontal Cortex* (Fourth Edition). London: Academic Press.

Gass, C. S. (1995). A procedure for assessing storage and retrieval on the Wechsler Memory Scale-Revised. *Archives of Clinical Neuropsychology,* 10(5), 475-487.

Gfeller, J.D., Meldrum, D.L., & Jacobi, K.A. (1995). The impact of constructional impairment on the WMS-R Visual Reproduction subtest. *Journal of Clinical Psychology,* 51:58-63.

Glosser, G., Goodglass, H., & Biber, C. (1989). Assessing visual memory disorders. *Psychological Assessment,* 1, 82 – 91.

Gold, J. M., Randolph, C., Carpenter, C. J., Goldberg, T.E., & Weinberger, D. R. (1992). The performance of patients with schizophrenia on the Wechsler Memory Scale-Revised. *The Clinical Neurologist,* 6, 367-373.

Goldman-Rakic, P.S. (1995). Architecture of the Prefrontal Cortex and the Central Executive. *Annals of the New York Academy of Sciences,* 769, 71–84.

Goodale, M.A., & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences,* 15(1), 20-25.

Graham, K. S., Simons, J. S., Pratt, K. H., Patterson, K., & Hodges, J. R. (2000). Insights from semantic dementia on the relationship between episodic and semantic dementia. *Neuropsychologia,* 38, 313 – 324.

Hanger, P.A., Montague, J.R., & Smith, M.C. (1991). A recognition memory test for the Visual Reproduction subtest of the. Wechsler Memory Scale-Revised. *Journal of Clinical and Experimental Neuropsychology,* 11, 605-614.

Haut, M.W., Weber, A.M., Demarest, D., Keefover, R.W., & Rankin, E.D. (1996). Controlling for constructional dysfunction with the visual reproduction subtest of the Wechsler Memory Scale-Revised in Alzheimer's Disease. *The Clinical Neuropsychologist,* 10(3), 309-312.

Haut, M.W., Weber, A.M., Wilhelm, K.L., Keefover, R.W., & Rankin, E.D. (1994). The visual reproduction subtest as a measure of visual perceptual/constructional functioning in dementia of the Alzheimer's type. *Clinical Neuropsychologist,* 8(2), 187-192.

Heilbronner, R. L. (1992). The search for a "pure" visual memory test: Pursuit of perfection. *The Clinical Neuropsychologist,* 6, 105-112.

Heilbronner, R.L., Buck, P., & Adams, R.L. (1989). Factor analysis of verbal and nonverbal clinical memory tests. *Archives of Clinical Neuropsychology*, 4(4), 299-309.

Helmstaedter, C, Pohl, C, Hufnagel, A., & Elger, C. E. (1991). Visual learning deficits in nonresected patients with right temporal lobe epilepsy. *Cortex*, 27, 547-555

Herman, B. P., Connell, B., Barr, W. B., & Wyler, A. R. (1995). The utility of the Warrington Recognition Memory Test for temporal lobe epilepsy: Pre and postoperative results. *Journal of Epilepsy,* 8, 139-145

Holdnack, J.A., Lissner, D., Bowden, S.C., & McCarthy, K.L. (2004) Utilizing the WAIS-III/WMS-III in Clinical Practice: Update of Research and Issues Relevant to Australian Normative Research. *Australian Psychologist*, 39(3), 220-227.

Howieson, D. B., & Lezak, M. D. (2004). Separating memory from other cognitive disorders. In A. D. Baddeley, M. D. Kopelman & B. A. Wilson (Eds.), *The Essential Handbook of Memory Disorders* (pp. 179 – 193). England: John Wiley & Sons.

Huguelet, P., Zanello, A., & Nicastro, R. (2000). A study of visual and auditory verbal working memory in schizophrenic patients compared to healthy subjects. *European Archives of Psychiatry and Clinical Neuroscience*, 250(2), 79-85.

Ivison, D. J. (1977). The Wechsler Memory Scale: preliminary findings towards an Australian standardization. *Australian Psychologist*, 12, 303-313.

Ivnik, R. J., Sharbrough, F. W., & Laws, E. (1987). Effects of anterior temporal lobectomy on cognitive function. *Journal of Clinical Psychology*, 43, 128-137.

Ivnik, R.J., Malec, J.F., Smith, G.E., Tangalos, E.G., Petersen, R.C., Kokmen, E., & Kurland, L.T. (1992). Mayo's older Americans normative studies: WMS-R norms for ages 56 to 94. *The Clinical Neuropsychologist*, 6(1), 49-82.

Jacobs, D.U., Troester, A.I., Butters, N., Salmon, D.P., & Cermak, L.S. (1990). Intrusion errors on the Visual Reproduction Test of the Wechsler Memory Scale and the Wechsler Memory Scale-Revised: An analysis of demented and amnesic patients. *Clinical Neuropsychologist*, 4(2), 177-191.

Jones, B., & Connolly, K. (1970). Memory effects in cross-modal matching. British *Journal of Psychology*, 61(2), 267-270.

Jones-Gotman, M. (1986). Right hippocampal excision impairs learning and recall of a list of abstract designs. *Neuropsychologia*, 24, 659 – 670.

Jurden, F.H., Franzen, M.D., Callahan, T., & Ledbetter, M. (1996). Factorial equivalence of the Wechsler Memory Scale-Revised across standardization and clinical samples. *Applied Neuropsychology*, 3, 65-74.

Kandel, E.R. (2004). The molecular biology of memory storage: a dialogue between genes and synapses. *Bioscience Reports*, 24(4-5), 475-522

Kapur, N. (1988). Memory Disorders in Clinical Practice. London: Butterworths. LaBar, K. S., & LeDoux, J. E. (2003). Emotion and the brain: An overview. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral Neurology and Neuropsychology* (2nd ed., pp. 711 – 724).

Kapur, S., Craik, F.I.M., Jones, C., Brown, G.M., Houle, S., & Tulving, E. (1995). Functional role of the prefrontal cortex in retrieval of memories: a PET study. *Neuroreport*, 6, 1880-1884.

Kelley, W.L., Miezin, F.M., McDermott, K., Buckner, R.L., Raichle, M.E., Cohen, N.J., & Petersen, S.E. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobes for verbal and nonverbal memory encoding. *Neuron,* 20, 927-936.

Knowlton, B.J., Mangels, J.A., & Squire, L.R. (1996). A neostriatal habit learning system in humans. *Science,* 273, 1399-1402.

Kopelman, M. D. (1995). The Korsakoff syndrome. *British Journal of Psychiatry,* 166, 154-173.

Larrabee, G. J., Kane, R. L., & Schuck, J. R. (1983). Factor analysis of the WAIS and Wechsler Memory Scale: an analysis of the construct validity of the Wechsler Memory Scale. *Journal of Clinical Neuropsychology,* 5, 159-168.

Larrabee, G. J., Kane, R. L., Schuck, J. R., & Francis, D. J. (1985). Construct validity of various memory testing procedures. *Journal of Clinical and Experimental Neuropsychology*, 7, 239-250.

Larrabee, G.J., & Curtiss, G. (1995). Construct validity of various verbal and visual memory tests. *Journal of Clinical and Experimental Psychology*, 17, 536-547.

LeDoux, J.E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience,* 23, 155-184.

Lee, G. P., Loring, D.W., & Thompson, J. L. (1989). Construct validity of material-specific memory measures following unilateral temporal lobe ablations. *Psychological Assessment: A Journal of Consulting and Clinical Psychology,* 1, 192-197.

Lee, T. M. C., Yip, J. T. H., & Jones-Gotman, M. (2002). Memory deficits after resection from left or right anterior temporal lobe in humans: A meta-analytic review. *Epilepsia,* 43, 283 – 291.

Levin, H. S., Mattis, S., Ruff, R. M., Eisenberg, H. M., Marshall, L. F., Tabaddor, K., High, W. M., & Frankowski, R. F. (1987). Neurobehavioral outcome following minor head injury: A three centre study. *Journal of Neurosurgery,* 66, 234-243.

Lewis, S.J.G., Dove, A., Robbins, T.W., Barker, R.A., & Owen, A.M. (2004). Striatal contributions to working memory: a functional magnetic resonance imaging study in humans. *European Journal of Neuroscience,* 19(3), 755-760.

Lezak, M. D. (1995). *Neuropsychological assessment-third edition.* New York: Oxford University Press.

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

Li, F., & Tsien, J.Z. (2009). Clinical Implications of Basic Research: Memory and the NMDA receptors. *New England Journal of Medicine,* 361, 302-303.

Loring, D.W. (1989) The Wechsler Memory Scale-Revised or the Wechsler Memory Scale-Revisited? *The Clinical Neuropsychologist,* 3:56-69.

Loring, D.W., & Papanicolaou, A.C. (1987). Memory assessment in neuropsychology: theoretical considerations and practical utility. *Journal of Clinical and Experimental Neuropsychology,* 9, 340-358.

Loring, D.W., Lee, G. P., Martin, R.C., & Meador, K.J. (1989). Verbal and visual memory index discrepancies from the Wechsler Memory Scale-Revised: cautions in interpretation. *Journal of Consulting and Clinical Psychology,* 1 (3), 198-202

Markowitsch, H. J. (2000). Neuroanatomy of memory. In E. Tulving., & Craik, F. I. M (Eds.), *The Oxford Handbook of Memory* (pp. 465 – 484). New York: Oxford University Press.

Markowitsch, H. J. (2003). Functional neuroanatomy of learning and memory. In P. W. Halligan., U. Kischka., & J. C. Marshall (Eds.), *Handbook of Clinical Neuropsychology* (pp. 724 – 740) New York: Oxford University Press.

Mayes, A. R. (1986). Learning and memory disorders and their assessment. *Neuropsychologia*, 24, 25 – 39.

Mayes, A. R. (1995). The assessment of memory disorders. In A. D. Baddeley, B. A. Wilson, & F. N. Watts (Eds.), *Handbook of Memory Disorders* (pp. 367 – 392). England: John Wiley & Sons.

Mayes, A. R., & Montaldi, D. (1999). The neuroimaging of long-term memory encoding processes. *Memory*, 7, 613 – 660.

Mayes, A. R., Meudell, P. R., Mann, D., & Pickering, A. (1988). Location of lesions in Korsakoff s syndrome: Neuropsychological and neuropathological data on two patients. *Cortex*, 24, 367-388.

Mayes, A., & Montaldi, D. (2001). Exploring the neural bases of episodic and semantic memory: the role of structural and functional neuroimaging. *Neuroscience and Biobehavioral Reviews*, 25, 555-573.

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field.* USA: Sage Publications.

McGaugh, J.L., Introini-Collison I.B., Cahill, L.F., Castellano, C., Dalmaz, C., Parent, M.B., & Williams, C.L. (1993). Neuromodulatory systems and memory storage: role of the amygdala. *Behavioural Brain Research*, 58(1-2), 81-90.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial Professional Manual.* Odessa: Psychological Assessment Resources Inc.

Milberg, W.P., Hebben, N., & Kaplan, E. (1986). The Boston process approach to neuropsychological assessment. In I. Grant & K.M. Adams (Eds.), *Neuropsychological Assessment of Neuropsychiatric Disorders* (pp. 56-86). New York: Oxford University Press.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Miller, L. A., Lai, R., & Munoz, D. G. (1998). Contributions of the entorhinal cortex, amygdala and hippocampus to human memory. *Neuropsychologia,* 36, 1247 – 1256.

Milner, B. (1965). Visually guided maze learning in man. Effects of bilateral hippocampus, bilateral frontal and bilateral unilateral cerebral lesions. *Neuropsychologia,* 3, 317-338.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin,* 27, 272 – 277.

Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not separate removal of amygdala and hippocampus. *Nature,* 273, 297-298.

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London*, B298, 85 – 92.

Mishkin, M., & Ungerleider, L.G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research,* 6 (1), 57-77.

Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: two neural systems. In: Lynch, G., McGaugh, J.L., Weinberger, N.M. (Eds.), *Neurobiology of Learning and Memory* (pp. 65-77). New York: Guilford.

Mishkin, M., Vargha-Khadem, F., & Gadian D. G. (1998). Amnesia and the organization of the hippocampal system. *Hippocampus*, 8, 212 – 216.

Mitrushina, M. N., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Mitrushina, M.N., Boone, K.B., & D'Elia, L.F. (1999). *Handbook of normative data for neuropsychological assessment.* New York: Oxford University Press.

Mittenberg, W., Burton, D. B., Darrow, E., & Thompson, G. B. (1992). Normative data for the Wechsler Memory Scale-Revised: Twenty-five-thirty-four year olds. *Psychological Assessment,* 4, 3363-3368.

Morris, R. G., Abrahams, S. & Polkey, C. E. (1995). Recognition memory for words and faces following unilateral temporal lobectomy. *British Journal of Clinical Psychology,* 34, 571-576

Moye, J. (1997). Nonverbal memory assessment with designs: Construct validity and clinical utility. *Neuropsychology Review,* 7, 157 – 170.

Nadel, L. (1994). Multiple memory systems: What and why, an update. In D. L. Schacter & E. Tulving (Eds.), *Memory Systems* (pp. 39 – 58). Cambridge, MA: MIT Press.

Nakashiba, T., Young, J.Z., McHugh, T.J., Buhl, D.L., & Tonegawa, S. (2008). Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science,* 319, 1260-1264.

Naugle, R.I., Chelune, G.J., Cheek, R., Luders, H., & Awad, I.A. (1993). Detection of changes in material-specific memory following temporal lobectomy using the Wechsler Memory Scale-Revised. *Archives of Clinical Neuropsychology*, 8, 381-395.

Panatier, A., Theodosis, D.T., Mothet, J.P., Touquet, B., Pollegioni, L., Poulain, D.A., & Oliet, S.H. (2006). Glia-derived D-serine controls NMDA receptor activity and synaptic memory. *Cell*, 125(4), 775-784.

Paniak, C., Murphy, D., Miller, H., & Lee, M. (1998). Wechsler Memory Scale-Revised Logical Memory and Visual Reproduction norms for 9- to 15-year-olds. *Developmental Neuropsychology,* 14(4), 555-562.

Pantel, J., Kratz, B., Essig, M., & Schröder, J. (2003). Parahippocampal volume deficits in subjects with aging-associated cognitive decline. *American Journal of Psychiatry,* 160(2), 379-382.

Papanicolaou, A. C. (2006). *The amnesias: A clinical textbook of memory disorders.* USA: Oxford University Press.

Parkin, A. J. (1997). *Memory and Amnesia: An introduction* (2nd ed.). Oxford: Blackwell Publishers Ltd.

Parkin, A.J., & Leng, N.R.C. (1993). *Neuropsychology of the Amnesic Syndrome*. East Sussex: Lawrence Erlbaum Associates.

Petersen, R.C., Smith, G., Kokmen, E., Ivnik, R. J., & Tangalos, E. G. (1992). Memory function in normal aging. *Neurology,* 42, 396-401.

Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology,* 58, 193-198.

Piguet, O., Saling, M. M., O'Shea, M. F., Berkovic, S. F., & Bladin, P. F. (1994). Rey figure distortions reflect nonverbal recall differences between right and left foci in unilateral temporal lobe epilepsy. *Archives of Clinical Neuropsychology,* 9, 451-460

Pliskin, N.H., Hamer, D.P., Goldstein, D.S., Towle, V.L., Reder, A.T., Noronha, A., & Arnason, B.G. (1996). Improved delayed visual reproduction test performance in multiple sclerosis patients receiving interferon beta-1b. *Neurology,* 47(6), 1463 - 1468.

Prado-Alcala, R.A., Ruiloba, M.I., Rubio, L., Solana-Figueroa, R., Medina, C., Salado-Castillo, R., & Quirarte, G.L. (2003). Regional infusions of serotonin into the striatum and memory consolidation. *Synapse,* 47(3), 169-175.

Prigatano, G.P. (1978). Wechsler Memory Scale: a selective review of the literature. *Journal of Clinical Psychology,* 34, 816-832.

Rausch, R., & Babb, T. L. (1993). Hippocampal neuron loss and memory scores before and after temporal lobe surgery for epilepsy. *Archives of Neurology,* 50, 812-817.

Reid, D. B., & Kelly, M. P. (1993). Wechsler Memory Scale-Revised in closed head injury. *Journal of Clinical Psychology*, 49, 245-254.

Rempel-Clower, N. L., Zola, S. M., Squire, L. R., & Amaral, D. G. (1996). Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. *Journal of Neuroscience,* 16, 5233 – 5255.

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

Russell, E.W. (1975). A multiple scoring method for the assessment of complex memory functions. *Journal of Consulting and Clinical Psychology*, 43, 800-809.

Ryan, J. J., Rosenberg, S. J., & Heilbronner, R. L. (1984). Comparative relationships of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and the Wechsler Adult Intelligence Scale (WAIS) to the Wechsler Memory Scale (WMS). *Journal of Behavioural Assessment*, 6, 37-43.

Ryan, J.J., & Lewis, C.V. (1988). Comparison of normal controls and recently detoxified alcoholics on the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist*, 2(2), 173-180.

Saling, M. M., Berkovic, S. F., O'Shea, M. F., Kalnins, R. M., Darby, D. G., & Bladin, P. F. (1993). Lateralisation of verbal memory and unilateral hippocampal sclerosis: Evidence of task-specific effects. *Journal of Clinical and Experimental Neuropsychology,* 15, 608-618.

Scharfman, H.E., Witter, M.P., & Schwarcz, R. (2000) *The Parahippocampal Region: Basic Science and Clinical Implications.* New York: New York Academy of Sciences.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, Psychiatry*, 20, 11 – 21.

Shepherd, G. M. (1994). Neurobiology, 3^ Edn. New York: Oxford University Press.

Shores, E.A., and Carstairs, J.R. (2000). The Macquarie University neuropsychological normative study (MUNNS): Australian norms for the WAIS-R and WMS-R. *Australian Psychologist,* 35(1), 41-59.

Smith, G. E., & Milner, B (1989). Right hippocampal impairment in the recall of spatial location: Encoding deficits or rapid forgetting? *Neuropsychologia,* 27, 71-81.

Smith, G. E., Malec, J. F., & Ivnik, R. J. (1992). Validity of the construct of nonverbal memory: A factor analytic study in a normal elderly sample. *Journal of Clinical and Experimental Neuropsychology,* 14, 211-221.

Smith, M. L., & Milner, B (1988). Estimation of frequency of occurrence of abstract designs after frontal or temporal lobectomy. *Neuropsychologia,* 26, 297-306.

Sperling, G. (1960). The information available in brief visual presentations. *Psychology Monographs,* 74(11), No. 498.

Spreen, O. and Strauss, E. (1998). *A Compendium of Neuropsychological Tests-second edition.* New York: Oxford University Press.

Squire L.R., & Zola-Morgan. S. (1991). The medial temporal lobe memory system. *Science,* 20, 1380-1386.

Squire, L. R. (1986). The neuropsychology of memory dysfunction and its assessment. In I. Grant., & K. M. Adams (Eds.), *Neuropsychological Assessment of Neuropsychiatric Disorders* (pp. 245 – 312). New York: Oxford University Press.

Squire, L. R., & Butters, N. (1992). *Neuropsychology of Memory* (2nd ed.). New York: Guilford Press.

Squire, L. R., & Shimamura, A. (1996). The neuropsychology of memory dysfunction and its assessment. In: Grant, I. & Adams, K. (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (2nd Ed., pp. 232-262). New York: Oxford University Press.

Squire, L. R., Amaral, D.G., &, Press, G. A. (1990). Magnetic resonance measurements of hippocampal formation and mammillary nuclei distinguish medial temporal lobe and diencephalic amnesia. *Journal of Neuroscience,* 10, 3106-3117.

Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review Psychology,* 44, 453 – 495.

Squire, L.R. & Zola-Morgan, S. (1988). Memory: Brain systems and behaviour. *Trends in Neuroscience,* 77,170-175.

Squire, L.R. (1987). *Memory and Brain.* New York: Oxford University Press.

Squire, L.R. (1992). Memory and the Hippocampus: A synthesis from findings with rats, monkeys and humans. *Psychological Review,* 99, 195-231.

Squire, L.R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory,* 82(3), 171-177.

Squire, L.R. (2009a). The legacy of patient H.M. for neuroscience. *Neuron,* 61, 6-9.

Squire, L.R. (2009b). Memory and brain systems: 1969-2009. *The Journal of Neuroscience,* 29(41), 12711-12716.

Squire, L.R., & Knowlton, B.J. (1994). Memory, hippocampus, and brain systems. In Gazzaniga, M. S. (Ed.), *The cognitive neurosciences* (pp. 825-837). Cambridge, Massachusetts: The Massachusetts Institute of Technology Press.

Squire, L.R., Stark, C.E.L, Clark, R.E. (2004). The medial temporal lobe. *Annual Review of Neurosciences,* 27, 279-306.

Suzuki, W.A., & Amaral, D.G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *The Journal of Comparative Neurology,* 350, 497-533.

Suzuki, W.A., & Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the New York Academy of Sciences,* 911, 175-191.

Thompson, R. F. (1986). The neurobiology of learning and memory. *Science,* 233, 941-947.

Todd, J.J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature,* 428, 751-754.

Trahan, D. E., Quintana, J. W., Willingham, A.C., & Goethe, K. E. (1988). The Visual Reproduction subtest: Standardisation and clinical validation of a delayed recall procedure. *Neuropsychology,* 2, 29-39.

Tranel, D. & Damasio, A. R. (1995). Neurobiological foundations of human memory. In A. D. Baddeley, B. A. Wilson & F. N. Watts (Eds.), *Handbook of memory disorders* (pp. 27-52). Chichester: John Wiley and Sons.

Trenerry, M. R., Jack, C. R., Cascino, G. D., Sharbrough, F. W. & Ivnik, R. J. (1996). Sex differences in the relationship between visual memory and MRI hippocampal volumes. *Neuropsychology,* 10, 343-351.

Troester, A. I. (1998). *Memory in neurodegenerative disease: Biological, cognitive, and clinical perspectives.* Cambridge, United Kingdom: Cambridge University Press.

Troester, A. I., Butters, N., Salmon, D. P., Cullum, C. M., Jacobs, D., Brandt, J., & White, R. F. (1993). The diagnostic utility of savings scores: Differentiating Alzheimer's and Huntington's diseases with logical memory and visual reproduction tests. *Journal of Clinical and Experimental Neuropsychology*, 15, 773-788.

Tsien, J.Z., Huerta, P.T., & Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell,* 87(7), 1327-1338.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of Memory.* New York: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26, 1-12.

Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology,* 53, 1-25.

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: Role of the hippocampus. *Hippocampus,* 8, 198 – 204.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review,* 80, 352 – 373.

Walker, A.J., Batchelor, J., & Shores, E.A. (2009). Effects of education and cultural background on performance on WAIS-III, WMS-III, WAIS-R and WMS-R measures: systematic review. *Australian Psychologist,* 44(4), 216-223.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology,* 19, 87-95.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.

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

Wilhelm, K.L. (1996). Developing a supplemental scoring system for the Wechsler Memory Scale-Revised. *Dissertation Abstracts International: Section B: The Sciences and Engineering,* 56(8-B), 4636.

Wilhelm, K.L., & Johnstone, B. (1995). Use of the Wechsler Memory Scale-Revised in traumatic brain injury. *Applied Neuropsychology,* 2(1), 42-45.

Williams, M.A., Rich, M.A., Reed, L.K., Jackson, W.T., LaMarche, J.A., & Boll, T.J. (1998). Visual reproduction subtest of the Wechsler Memory Scale-Revised: analysis of construct validity. *Journal of Clinical Psychology,* 54(7), 963-971.

Wilson, B. A. (2004). Assessment of memory disorders. In A. D. Baddeley., M. D. Kopelman., & B. A. Wilson (Eds.), *The Essential Handbook of Memory Disorders* (pp. 159 – 178). England: John Wiley & Sons.

Winocur, G., Moscovitch, M., Caruana, D. A., & Binns, M. A. (2005). Retrograde amnesia in rats with lesions to the hippocampus on a test of spatial memory. *Neuropsychologia,* 43, 1580 – 1590.

Woloszyn, D. B., Grob-Murphy, S., Wetzel, L., & Fisher, W. (1993). Inter-rater agreement on the Wechsler Memory Scale-Revised in a mixed clinical population. *The Clinical Neuropsychologist,* 7, 467-471.

Wong, J.L., & Gilpin, A.R. (1993). Verbal vs. visual categories on the Wechsler Memory Scale-Revised: How meaningful a distinction? *Journal of Clinical Psychology,* 49, 847-854.

Zola, M. S. (1998). Memory, amnesia, and issue of recovered memory: Neurobiological aspects. *Clinical Psychology Review,* 18, 915 – 932.

Zola, S. (1997). Amnesia: Neuroanatomic and clinical aspects. In T. E. Feinberg and M. J. Farah (Eds.), *Behavioural neurology and neuropsychology* (pp. 447-461). New York: McGraw Hill

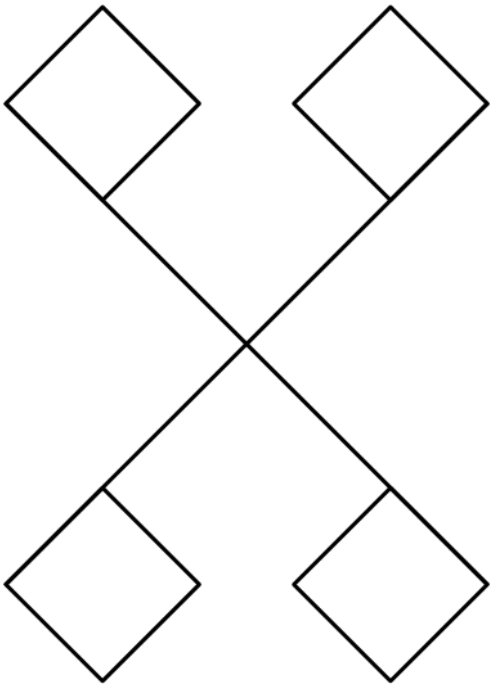
Zola-Morgan, S., & Squire, L.R. (1993). Neuroanatomy of Memory. *Annual Review of Neuroscience,* 16, 547-563.

Zola-Morgan, S., Squire, L. R. & Amaral, D. (1986). Human amnesia and the medial temporal region: Enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *Journal of Neuroscience,* 6, 2950 – 2 967.

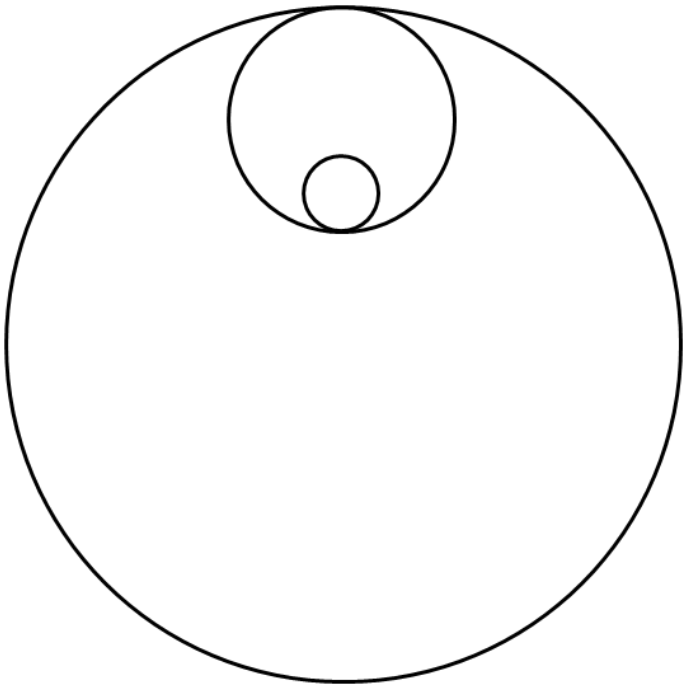
Zola-Morgan, S., Squire, L.R., Amaral, D.G., & Suzuki, W.A. (1989). Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *Journal of Neuroscience,* 9, 4355-4370.

Appendix A.

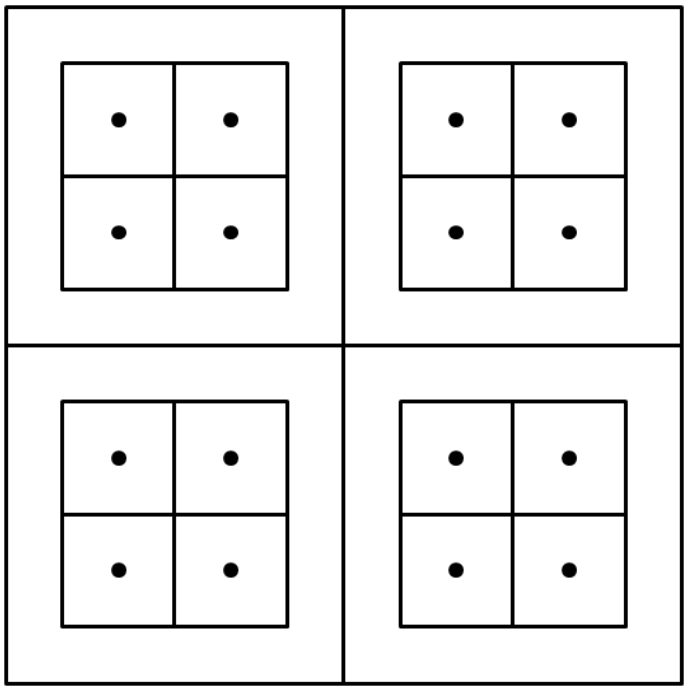
Stimulus cards from the original WMS-R



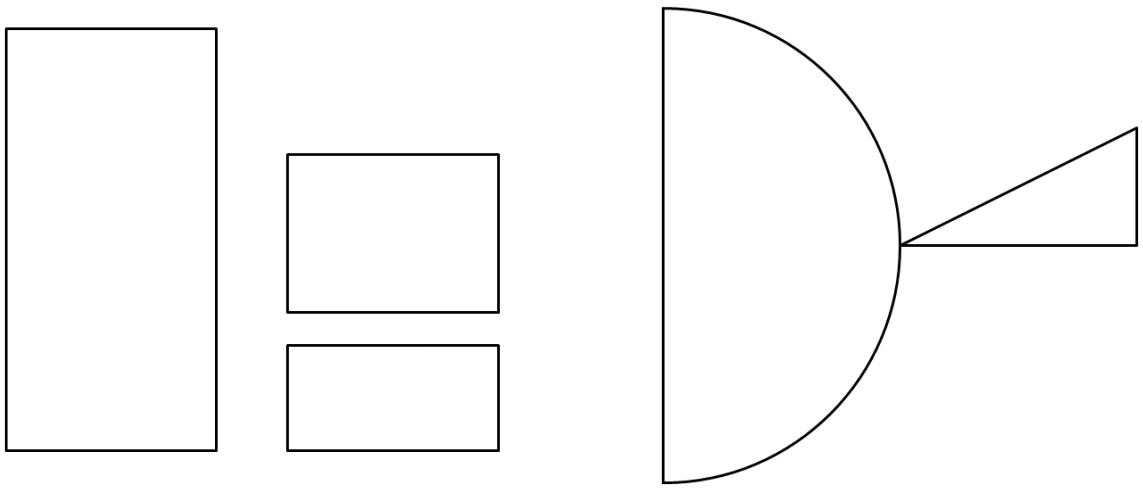
Card A



Card B



Card C



Card D

Appendix B.

Alternate Scoring System -Wording Changes to Scoring Criteria

Design	Item	Description
1	9.	All figures are identical in shape and size to each other (90% ~~tolerance~~ the same).
1	11.	At least ¾ of the figures ~~are squares (flags) i.e. they~~ have four sides. Gaps between the line and figure are acceptable if no greater than 25% of the length of the side of the figure where gap is present.
2	12.	The medium figure is ~~about 1/3~~ 25-50% the diameter of the vertical axis of the large figure. ~~Acceptable range is 25%-50%~~
2	13.	The small figure is ~~about 1/3~~ 25-50% the diameter of the vertical axis of the medium figure. ~~Acceptable range is 25%-50%~~
2	14.	~~Areas enclosed by each figure are in correct relative proportion.~~ The area of the small figure is about 25-50% of the area of the medium figure, and the area of the medium figure is about 25-50% of the area of the large figure.
2	18.	At least two of the figures are discrete circles rather than ovals (i.e. ~~separate circles in their own right). For each circle,~~ the smallest diameter is at least 90% of the largest diameter). ~~Any common circumference is <20% of the circumference of the larger of the two figures.~~
2	19.	All figures are discrete circles rather than ovals ~~or any other shape. For each circle,~~ (i.e. the smallest diameter is at least 90% of the largest diameter). ~~Any common circumference is <20% of the circumference of the larger of the two figures.~~
3	4.	On the four-sided figure ~~is exactly a square,~~ every angle is in the range 85-95˚. The smallest line is at least 90% of the largest line. Gaps or overlaps are acceptable as long as they are less than 10% of the length of the line.
3	10.	Smaller figures ~~are distinct shapes (even if there are less than or more than four figures). The figures~~ do not overlap each other or the sides of the square or the internal divisions or any additional lines. They do not share a common border, but may touch or partially overlap.

* Wording deleted is indicated in ~~strike-though~~, wording added is indicated in bold type

Table continued on following page.

3	13.	The smaller figures (as drawn by the ~~client~~ subject) are ~~similar (~~90%~~)~~ the same in configuration or size.
3	18.	The dot or open circle occupies less than 10% of the area in each segment of the internal figure, ~~in each figure if no segments, or in each quadrant if no smaller figures.~~
4	7.	The base of the tall rectangle and the base of the lowest adjacent figure ~~are level~~ must be within 10% of the height of the tall rectangle. ~~The adjacent figure(s) need not be four-sided.~~
4	13.	~~The semicircle is in the correct proportion.~~ The radius of the semicircle is ~~half~~ 40-60% the size of the vertical dimension. ~~(Acceptable range is 40%-60% of vertical dimension).~~
4	14.	The ~~figure as described in Item 10~~ semicircle is located to the right of ~~the figures mentioned in Items 1-9, or to the right of~~ some other shape or even a line. If there is nothing to the left of the ~~figure in Item 10~~ semicircle, score zero.

* Wording deleted is indicated in ~~strike-though~~, wording added is indicated in bold type

Appendix C.

Alternate Scoring System - Refined scoring items for Design 1

Item	Description
1.	There are at least two continuous lines or four lines emanating from a central point or figure. If there is only one line or more than four lines from a central point, score zero.
2.	There are only two lines present and these two lines intersect. If there are more or less than two lines, score zero.
3.	Two lines intersect in the middle 1/3 of each other. A minor gap at or near the intersection is acceptable if there is no change of direction. If the figure has a radial spokes design, all spokes are of the same size (The shortest spoke must be at least 50% of the longest spoke).
4.	The lines that intersect or emanate from the central point do not form angles less than 45˚.
5.	Lines or radial spokes are similar in length. The shorter line or spoke must be at least 75% of the longest line or spoke (regardless of where they intersect).
6.	The intersecting lines have not been rotated to an orthogonal position. If one line is vertical, the other is not horizontal or, if one line is horizontal, the other is not vertical. If the lines do not intersect, score zero.
7.	At least three geometric figures are present.
8.	All figures have the same number of sides (or all are circular). Figures can share a side with the lines/spokes, but they cannot share a border with an external figure (e.g. a bordering square).
9.	* All figures are identical in shape and size to each other (90% the same).
10.	All figures are between 30-50% of the length of the radial arm of the line. If the figure/s is not joined to the line, use the longest side of the figure as the reference comparison to the radial arm. If there are no lines, score zero.
11.	* At least ¾ of the figures have four sides. Gaps between the line and figure are acceptable if no greater than 25% of the length of the side of the figure where gap is present.
12.	Exactly four discrete figures are present. The figures do not share a border with an external square.

* Items that underwent revision for the present study are marked with an asterisk (*).

Table continued on following page.

13.	More than ½ of the figures touch any line emanating from the central point. Figures may overlap the line.
14.	At least two figures are correctly positioned near the end of the appropriate line. A figure does not have to touch or overlap the adjacent line.
15.	All four figures are correctly positioned near the end of the appropriate line. A figure does not have to touch or overlap the adjacent line.
16.	All four figures touch one line at the endpoint of the line. Minor overlap or gaps (less than 10% of the longest side of the figure) are acceptable. If the figure/s is a circle, overshoot or gap must be less than 10% of the diameter of the circle.
17.	The side of all four figure/s is contiguous with the line (i.e. the figure/s share a side with the line). A minor gap in the line before it joins the side of the figure is acceptable (less than 10% of the length of the side of the figure).
18.	Two figures face inwards (if rotation of the lines is less than 90˚, assume direction that would maximize the score). If any figure is not contiguous with the line and 100% of that figure is not on the correct side of the adjacent line, score zero.
19.	Four figures face inwards. If any figure is not contiguous with the line and 100% of that figure is not on the correct side of the adjacent line, score zero.
20.	No extra elements are present. Minor overshoots of lines should not be penalized.

* Items that underwent revision for the present study are marked with an asterisk (*).

Appendix D.

Alternate Scoring System - Refined scoring items for Design 2

Item	Description
1.	At least one circular figure is present.
2.	Three geometric figures (only) are present and at least one is circular in shape.
3.	At least two geometric figures are present, with one mostly inside the other. Figures may share a border.
4.	The figures form a clear size gradient (i.e. they are not of equal size). If there are more than three figures, take the largest one to be the large figure and the smallest one to be the small figure, then choose the medium figure so as to maximize the score. If there are only two figures, they must form a clear size gradient. A dot is not a figure.
	If there are only two figures, interpret spatial relationships questions (Items 5 – 14) so as to maximize the score.
5.	The large figure mostly encloses at least one smaller figure.
6.	The large figure mostly encloses two smaller figures.
7.	The small figure is mostly enclosed by a medium figure.
8.	A medium figure is located towards the top of the large figure and away from the bottom. The gap between the bottom of the large figure and the bottom of the medium figure should be at least three times the size of the gap between the top of those figures.
9.	The top of a medium figure touches the top of the larger figure. Minor overlap or gap between the figures is acceptable (less than 10% the diameter of the large figure).
10.	A small figure is located towards the bottom of the medium figure and away from the top (regardless of whether it is enclosed by the medium figure). The gap between the top of the medium figure and the top of the small figure should be at least three times the gap between the bottom those figures.
11.	The bottom of the small figure touches the bottom of the medium figure. Minor overlap or gap between the figures is acceptable (less than 10% the diameter of the medium figure).

* Items that underwent revision for the present study are marked with an asterisk (*).

Table continued on following page.

12.	* The medium figure is 25-50% the diameter of the vertical axis of the large figure.
13.	* The small figure is 25-50% the diameter of the vertical axis of the medium figure.
14.	* The area of the small figure is about 25-50% of the area of the medium figure, and the area of the medium figure is about 25-50% of the area of the large figure.
15.	The figures are symmetrically placed about the midline. If a vertical midline axis is drawn to divide the largest figure, no more than 60% of any figure is present on one side of that axis.
16.	The spatial relationship between the three figures is preserved, even if the relationship is inverted. If there are only two circles, score zero.
17.	All figures are primarily closed and circular (can be ovals). Any common circumference is less than 20% of the circumference of the larger of the two figures.
18.	* At least two of the figures are discrete circles rather than ovals (i.e. the smallest diameter is at least 90% of the largest diameter).
19.	* All figures are discrete circles rather than ovals (i.e. the smallest diameter is at least 90% of the largest diameter).
20.	No extra elements are present, except minor line continuations.

* Items that underwent revision for the present study are marked with an asterisk (*).

Appendix E.

Alternate Scoring System - Refined scoring items for Design 3

Item	Description
1.	A large figure with two or more internal elements (lines, figures) is present. The large figure may share a side with the edge of the paper for this item only. If in any doubt, interpret to maximize score. If two outer squares, consider the outermost to be the large square.
2.	At least one large four-sided figure is present and it is approximately square. The figure may be rectangular as long as the shorter side is at least 50% of the length of the longer sides.
3.	The four sides of the large square are reasonably equal in length. The longest side is no more than 25% longer than the shortest side.
4.	* On the four-sided figure, every angle is in the range 85-95˚. The smallest line is at least 90% of the largest line. Gaps or overlaps are acceptable as long as they are less than 10% of the length of the line.
5.	A vertical division divides the large figure. A double-lined vertical division is acceptable. The division can be contiguous with the internal squares. Gaps in the joining of the division and the external figure are acceptable as long as the length of the vertical division is at least 75% the length of the vertical dimension of the square.
6.	A horizontal division divides the large figure. A double-lined horizontal division is acceptable. The division can be contiguous with the internal squares. Gaps in the joining of the division and the external figure are acceptable as long as the length of the horizontal division is at least 75% the length of the horizontal dimension of the square.
7.	The vertical and horizontal divisions intersect and divide the figure into four quadrants (i.e. they touch and cross each other).
8.	Two to four smaller figures are present, with or without a major figure bordering them. Each smaller figure shares no more than two sides with any of the following: the horizontal division, the vertical division, the external square, or any other smaller figure.

* Items that underwent revision for the present study are marked with an asterisk (*).

Table continued on following page.

9.	Each quadrant of the larger figure has only one smaller figure. Quadrants need not be symmetrical. No more than two figures share a line with each other OR if there is no larger figure, the smaller figures form a 2x2 matrix and no more than two figures share a line.
10.	* Smaller figures do not overlap each other or the sides of the square or the internal divisions or any additional lines. They do not share a common border, but may touch or partially overlap.
11.	Each of the smaller figures is divided into four parts, or there are four shapes in each quadrant in a 2x2 matrix. The entire quadrant being divided into four scores zero.
12.	Each smaller figure is divided into four, or each quadrant is divided into four by a vertical and a horizontal line. Double vertical or horizontal lines are acceptable.
13.	* The smaller figures (as drawn by the subject) are 90% the same in configuration or size.
14.	At least three of the smaller figures are in correct proportion to the larger figure, as per the original design. If there is no large square, score zero.
15.	The smaller figures (as drawn by the subject) have four sides and are separate from each other, from the internal divisions and the external square. There is no overlap between sides.
16.	A number of dots (or circles) are present in at least 75% of the internal segments of the smaller figures, in at least 75% of each smaller figure if there are no segments, in at least 75% of the smaller figures that the subject produces; or in at least 75% of the quadrants.
17.	Each quadrant of the large figure has only four dots/circles in a square array. Divisions may or may not be present. If no larger figure, score zero.
18.	* The dot or open circle occupies less than 10% of the area in each segment of the internal figure.
19.	All four smaller figures are placed symmetrically. Borders are equal and less than 20% of the length of the quadrant. If no internal divisions are present or there is no external square, equal spacing occurs between the smaller figures in both the horizontal and vertical planes.
20.	No extra lines, dots or figures. Minor overshoots of lines should not be penalized.

* Items that underwent revision for the present study are marked with an asterisk (*).

Appendix F.

Alternate Scoring System - Refined scoring items for Design 4

Item	Description
1.	At least two figures are present, of which one is a four-sided or is a circle/ semicircle. Figures may share a common border. If there is one figure, score zero.
2.	A tall rectangle is present. The base of the rectangle is less than 75% of the vertical dimension. The longest side is no more than 20% longer the parallel side.
3.	One or more 3-6 sided figure/s is adjacent to the large rectangle (sharing a border is acceptable) OR there are one or more 3-6 sided figure/s (if no large rectangle is present).
4.	The smaller figure/s in Item 3 are separate from each other and from the major figure (rectangle). Minor touching or overlap is acceptable.
5.	The bases of all figures are of similar length (the smallest base is at least 90% of the largest base). If there is only one figure, score zero.
6.	The tall rectangle is clearly above the height of the adjacent figure/s by at least 10% of its height. The adjacent figure/s do not need to have four sides.
7.	* The base of the tall rectangle and the base of the lowest adjacent figure must be within 10% of the height of the tall rectangle.
8.	There are two four-sided figures positioned on top of each other and to the right of the large rectangle (if it is present). The two figures' widths are greater than their heights.
9.	Of the two four-sided figures in Item 8, one is clear larger and placed above the smaller figure. The smaller figure is no more than 70% the height of the figure above it, at any point.
10.	A large figure with a curved surface OR a curved line is present.
11.	The large figure in Item 10 is a discrete semicircle only (irrespective of orientation).
12.	The curved portion of the semicircle faces the right.
13.	* The radius of the semicircle is 40-60% the size of the vertical dimension.
14.	* The semicircle is located to the right of some other shape or even a line. If there is nothing to the left of the semicircle, score zero.

* Items that underwent revision for the present study are marked with an asterisk (*).

Table continued on following page.

15.	A smaller figure is located near the figure described in Item 10 OR if the figure described in Item 10 is absent, a smaller figure is placed to the far right of any other shapes (as per Items 1-9). It is acceptable if the smaller figure is placed inside the figure from Item 10. A line receives credit for this item but it would score zero for Items 16-19.
16.	The smaller figure in Item 15 is separate from any other figure. The smaller figure is not inside another figure and it does not share a side with another figure. The smaller figure can touch, overlap (10% of diameter tolerance) or be in close proximity to the figure described in Item 10.
17.	The smaller figure in Item 15 is located to the right of the figure described in Item 10, or to the right of a four-sided figure if the figure in Item 10 is absent.
18.	The smaller figure in Item 15 is located at or near the centre of the right border of the figure in Item 10. The smaller figure must be within 30˚ above or below the centre of the arc of the figure in Item 10 and the smaller figure can be inside that figure OR if the figure in Item 10 is absent, the smaller figure must be located above the level of the bases of the figures described in Items 1-9 and below the upper level of the edge of the large rectangle.
19.	The smaller figure in Item 15 is a discrete triangle (i.e. has three discrete sides separate from the figure described in Item 10).
20.	No extra elements are present. Minor overshoots of lines should not be penalized.

* Items that underwent revision for the present study are marked with an asterisk (*).

Appendix G.

Intra-rater reliability analysis

Immediate Recall

	Intraclass Correlations		
Design	Alternate Scoring System	Original Scoring System	*N*
1	.95	.94	40
2	.98	.91	40
3	.98	.96	40
4	.99	.93	40
Total	.99	.97	40

Delayed Recall

	Intraclass Correlations		
Design	Alternate Scoring System	Original Scoring System	*N*
1	.97	.98	40
2	.99	.97	40
3	.98	.97	40
4	.96	.97	40
Total	.99	.98	40

Appendix H.

Inter-rater reliability analysis

Immediate Recall

	Intraclass Correlations		
Design	Alternate Scoring System	Original Scoring System	*N*
1	.96	.83	40
2	.92	.89	40
3	.98	.82	40
4	.99	.93	40
Total	.98	.94	40

Delayed Recall

	Intraclass Correlations		
Design	Alternate Scoring System	Original Scoring System	*N*
1	.95	.81	40
2	.95	.85	40
3	.95	.91	40
4	.97	.87	40
Total	.98	.96	40

Appendix I.

Descriptive Statistics for the Original and Alternate Scoring Systems

Age Group in Years

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18-21 years	112	28.1	28.1	28.1
	22-25 years	89	22.3	22.3	50.4
	26-29 years	89	22.3	22.3	72.7
	30-34 years	109	27.3	27.3	100.0
	Total	399	100.0	100.0	

Education Group in Years

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8-12 years of education	182	45.6	45.6	45.6
	13-19 years of education	217	54.4	54.4	100.0
	Total	399	100.0	100.0	

Descriptive Statistics

	N	Mean	Std.	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Sex	399	1.52	.500	.250	-.065	.122	-2.006	.244
Age in years	399	25.64	4.973	24.734	.131	.122	-1.245	.244
Education in years	399	12.93	2.052	4.211	.123	.122	-.334	.244
WAIS-R: Verbal IQ	399	101.38	12.350	152.518	.262	.122	-.010	.244
WAIS-R: Performance IQ	399	102.67	12.528	156.939	-.081	.122	-.288	.244
WAIS-R: Full Scale IQ	399	101.93	12.358	152.719	.126	.122	-.170	.244
Valid N (listwise)	399							

Descriptive Statistics

	N	Mean	Std.	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
OS_IR_D1	399	5.7995	.85941	.739	-.819	.122	1.416	.244
OS_IR_D2	399	5.4837	1.28740	1.657	-.397	.122	-.718	.244
OS_IR_D3	399	7.3559	1.31235	1.722	-1.336	.122	2.501	.244
OS_IR_D4	399	16.1905	1.82358	3.325	-1.549	.122	2.927	.244
OS_IR_TO	399	34.8221	3.35443	11.252	-1.097	.122	2.192	.244
OS_DR_D1	399	4.9048	2.15270	4.634	-1.577	.122	1.113	.244
OS_DR_D2	399	4.5840	2.05266	4.213	-.946	.122	.267	.244
OS_DR_D3	399	7.0652	1.61207	2.599	-1.730	.122	4.208	.244
OS_DR_D4	399	15.4937	2.65700	7.060	-2.342	.122	8.084	.244
OS_DR_TO	399	32.0476	5.16270	26.654	-1.068	.122	2.022	.244
AS_IR_D1	399	18.1479	.90819	.825	-.033	.122	.215	.244
AS_IR_D2	399	17.9900	1.41595	2.005	-.601	.122	.344	.244
AS_IR_D3	399	17.1228	1.88287	3.545	-1.114	.122	2.457	.244
AS_IR_D4	399	18.1253	1.47145	2.165	-1.269	.122	3.437	.244
AS_IR_TO	399	71.3860	3.78789	14.348	-.653	.122	1.227	.244
AS_DR_D1	399	15.4286	6.36568	40.522	-1.974	.122	2.034	.244
AS_DR_D2	399	15.7118	5.68120	32.276	-2.215	.122	3.459	.244
AS_DR_D3	399	16.6341	2.79113	7.790	-3.147	.122	15.060	.244
AS_DR_D4	399	17.6140	2.34819	5.514	-3.431	.122	18.803	.244
AS_DR_TO	399	65.3885	9.73658	94.801	-1.086	.122	.811	.244
Valid N (listwise)	399							

Appendix J.

Percentile Ranks for Alternate Scoring System

AS_IR_D1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15.00	1	.3	.3	.3
	16.00	12	3.0	3.0	3.3
	17.00	68	17.0	17.0	20.3
	18.00	193	48.4	48.4	68.7
	19.00	96	24.1	24.1	92.7
	20.00	29	7.3	7.3	100.0
	Total	399	100.0	100.0	

AS_IR_D2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	12.00	1	.3	.3	.3
	14.00	5	1.3	1.3	1.5
	15.00	12	3.0	3.0	4.5
	16.00	42	10.5	10.5	15.0
	17.00	73	18.3	18.3	33.3
	18.00	110	27.6	27.6	60.9
	19.00	97	24.3	24.3	85.2
	20.00	59	14.8	14.8	100.0
	Total	399	100.0	100.0	

AS_IR_D3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8.00	1	.3	.3	.3
	9.00	1	.3	.3	.5
	11.00	4	1.0	1.0	1.5
	12.00	6	1.5	1.5	3.0
	13.00	4	1.0	1.0	4.0
	14.00	12	3.0	3.0	7.0
	15.00	37	9.3	9.3	16.3
	16.00	59	14.8	14.8	31.1
	17.00	87	21.8	21.8	52.9
	18.00	101	25.3	25.3	78.2
	19.00	57	14.3	14.3	92.5
	20.00	30	7.5	7.5	100.0
	Total	399	100.0	100.0	

AS_IR_D4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	11.00	3	.8	.8	.8
	13.00	1	.3	.3	1.0
	14.00	2	.5	.5	1.5
	15.00	15	3.8	3.8	5.3
	16.00	21	5.3	5.3	10.5
	17.00	69	17.3	17.3	27.8
	18.00	116	29.1	29.1	56.9
	19.00	104	26.1	26.1	83.0
	20.00	68	17.0	17.0	100.0
	Total	399	100.0	100.0	

AS_IR_TO

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	56.00	1	.3	.3	.3
	58.00	2	.5	.5	.8
	59.00	1	.3	.3	1.0
	61.00	2	.5	.5	1.5
	62.00	3	.8	.8	2.3
	63.00	6	1.5	1.5	3.8
	64.00	7	1.8	1.8	5.5
	65.00	3	.8	.8	6.3
	66.00	9	2.3	2.3	8.5
	67.00	22	5.5	5.5	14.0
	68.00	19	4.8	4.8	18.8
	69.00	29	7.3	7.3	26.1
	70.00	43	10.8	10.8	36.8
	71.00	41	10.3	10.3	47.1
	72.00	52	13.0	13.0	60.2
	73.00	43	10.8	10.8	70.9
	74.00	38	9.5	9.5	80.5
	75.00	35	8.8	8.8	89.2
	76.00	17	4.3	4.3	93.5
	77.00	14	3.5	3.5	97.0
	78.00	2	.5	.5	97.5
	79.00	5	1.3	1.3	98.7
	80.00	5	1.3	1.3	100.0
	Total	399	100.0	100.0	

AS_DR_D1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	57	14.3	14.3	14.3
	11.00	1	.3	.3	14.5
	15.00	4	1.0	1.0	15.5
	16.00	11	2.8	2.8	18.3
	17.00	50	12.5	12.5	30.8
	18.00	199	49.9	49.9	80.7
	19.00	63	15.8	15.8	96.5
	20.00	14	3.5	3.5	100.0
	Total	399	100.0	100.0	

AS_DR_D2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	43	10.8	10.8	10.8
	8.00	1	.3	.3	11.0
	12.00	3	.8	.8	11.8
	13.00	1	.3	.3	12.0
	14.00	9	2.3	2.3	14.3
	15.00	25	6.3	6.3	20.6
	16.00	31	7.8	7.8	28.3
	17.00	74	18.5	18.5	46.9
	18.00	102	25.6	25.6	72.4
	19.00	79	19.8	19.8	92.2
	20.00	31	7.8	7.8	100.0
	Total	399	100.0	100.0	

AS_DR_D3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	5	1.3	1.3	1.3
	7.00	2	.5	.5	1.8
	8.00	2	.5	.5	2.3
	10.00	4	1.0	1.0	3.3
	11.00	3	.8	.8	4.0
	12.00	3	.8	.8	4.8
	13.00	7	1.8	1.8	6.5
	14.00	12	3.0	3.0	9.5
	15.00	46	11.5	11.5	21.1
	16.00	67	16.8	16.8	37.8
	17.00	82	20.6	20.6	58.4
	18.00	94	23.6	23.6	82.0
	19.00	49	12.3	12.3	94.2
	20.00	23	5.8	5.8	100.0
	Total	399	100.0	100.0	

AS_DR_D4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	2	.5	.5	.5
	6.00	1	.3	.3	.8
	7.00	2	.5	.5	1.3
	10.00	3	.8	.8	2.0
	11.00	4	1.0	1.0	3.0
	12.00	1	.3	.3	3.3
	13.00	2	.5	.5	3.8
	14.00	10	2.5	2.5	6.3
	15.00	11	2.8	2.8	9.0
	16.00	27	6.8	6.8	15.8
	17.00	82	20.6	20.6	36.3
	18.00	113	28.3	28.3	64.7
	19.00	89	22.3	22.3	87.0
	20.00	52	13.0	13.0	100.0
	Total	399	100.0	100.0	

Appendix K.

Tests of Normality for the Original and Alternate Scoring Systems

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
OS_IR_D1	.286	399	.000	.844	399	.000
OS_IR_D2	.181	399	.000	.889	399	.000
OS_IR_D3	.240	399	.000	.852	399	.000
OS_IR_D4	.225	399	.000	.832	399	.000
OS_IR_TO	.149	399	.000	.934	399	.000
OS_DR_D1	.327	399	.000	.687	399	.000
OS_DR_D2	.197	399	.000	.863	399	.000
OS_DR_D3	.231	399	.000	.824	399	.000
OS_DR_D4	.215	399	.000	.776	399	.000
OS_DR_TO	.121	399	.000	.939	399	.000
AS_IR_D1	.251	399	.000	.890	399	.000
AS_IR_D2	.169	399	.000	.926	399	.000
AS_IR_D3	.163	399	.000	.914	399	.000
AS_IR_D4	.188	399	.000	.879	399	.000
AS_IR_TO	.097	399	.000	.968	399	.000
AS_DR_D1	.415	399	.000	.524	399	.000
AS_DR_D2	.315	399	.000	.590	399	.000
AS_DR_D3	.200	399	.000	.723	399	.000
AS_DR_D4	.239	399	.000	.701	399	.000
AS_DR_TO	.197	399	.000	.878	399	.000

a. Lilliefors Significance Correction

Appendix L.

Nonparametric Correlations

Correlations

			OS_IR_D1	OS_IR_D2	OS_IR_D3	OS_IR_D4	OS_IR_TO	OS_DR_D1	OS_DR_D2	OS_DR_D3	OS_DR_D4	OS_DR_TO
Spearman's rho	OS_IR_D1	Correlation Coefficient	1.000	.119*	.149**	.074	.404**	.316**	.140**	.156**	.161**	.259**
		Sig. (2-tailed)	.	.017	.003	.141	.000	.000	.005	.002	.001	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_D2	Correlation Coefficient	.119*	1.000	.193**	.151**	.602**	.049	.407**	.188**	.176**	.273**
		Sig. (2-tailed)	.017	.	.000	.002	.000	.332	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_D3	Correlation Coefficient	.149**	.193**	1.000	.266**	.623**	.135**	.201**	.574**	.263**	.410**
		Sig. (2-tailed)	.003	.000	.	.000	.000	.007	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_D4	Correlation Coefficient	.074	.151**	.266**	1.000	.679**	.118*	.157**	.293**	.676**	.500**
		Sig. (2-tailed)	.141	.002	.000	.	.000	.019	.002	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_TO	Correlation Coefficient	.404**	.602**	.623**	.679**	1.000	.199**	.377**	.480**	.573**	.608**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_D1	Correlation Coefficient	.316**	.049	.135**	.118*	.199**	1.000	.083	.209**	.111*	.528**
		Sig. (2-tailed)	.000	.332	.007	.019	.000	.	.098	.000	.027	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_D2	Correlation Coefficient	.140**	.407**	.201**	.157**	.377**	.083	1.000	.176**	.222**	.589**
		Sig. (2-tailed)	.005	.000	.000	.002	.000	.098	.	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_D3	Correlation Coefficient	.156**	.188**	.574**	.293**	.480**	.209**	.176**	1.000	.316**	.567**
		Sig. (2-tailed)	.002	.000	.000	.000	.000	.000	.000	.	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_D4	Correlation Coefficient	.161**	.176**	.263**	.676**	.573**	.111*	.222**	.316**	1.000	.659**
		Sig. (2-tailed)	.001	.000	.000	.000	.000	.027	.000	.000	.	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_TO	Correlation Coefficient	.259**	.273**	.410**	.500**	.608**	.528**	.589**	.567**	.659**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.
		N	399	399	399	399	399	399	399	399	399	399

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

			AS_IR_D1	AS_IR_D2	AS_IR_D3	AS_IR_D4	AS_IR_TO	AS_DR_D1	AS_DR_D2	AS_DR_D3	AS_DR_D4	AS_DR_TO
Spearman's rho	AS_IR_D1	Correlation Coefficient	1.000	.226**	.220**	.307**	.556**	.365**	.196**	.283**	.279**	.301**
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_IR_D2	Correlation Coefficient	.226**	1.000	.258**	.156**	.630**	.136**	.552**	.272**	.198**	.319**
		Sig. (2-tailed)	.000	.	.000	.002	.000	.007	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_IR_D3	Correlation Coefficient	.220**	.258**	1.000	.219**	.720**	.265**	.212**	.739**	.273**	.439**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_IR_D4	Correlation Coefficient	.307**	.156**	.219**	1.000	.599**	.148**	.102*	.241**	.677**	.324**
		Sig. (2-tailed)	.000	.002	.000	.	.000	.003	.041	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_IR_TO	Correlation Coefficient	.556**	.630**	.720**	.599**	1.000	.331**	.413**	.638**	.523**	.555**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_D1	Correlation Coefficient	.365**	.136**	.265**	.148**	.331**	1.000	.121*	.279**	.203**	.662**
		Sig. (2-tailed)	.000	.007	.000	.003	.000	.	.015	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_D2	Correlation Coefficient	.196**	.552**	.212**	.102*	.413**	.121*	1.000	.237**	.183**	.542**
		Sig. (2-tailed)	.000	.000	.000	.041	.000	.015	.	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_D3	Correlation Coefficient	.283**	.272**	.739**	.241**	.638**	.279**	.237**	1.000	.280**	.545**
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_D4	Correlation Coefficient	.279**	.198**	.273**	.677**	.523**	.203**	.183**	.280**	1.000	.498**
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_TO	Correlation Coefficient	.301**	.319**	.439**	.324**	.555**	.662**	.542**	.545**	.498**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.
		N	399	399	399	399	399	399	399	399	399	399

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

			AS_IR_D1	AS_IR_D2	AS_IR_D3	AS_IR_D4	AS_IR_TO	OS_IR_D1	OS_IR_D2	OS_IR_D3	OS_IR_D4	OS_IR_TO
Spearman's rho	AS_IR_D1	Correlation Coefficient	1.000	.226**	.220**	.307**	.556**	.507**	.238**	.178**	.241**	.436**
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_IR_D2	Correlation Coefficient	.226**	1.000	.258**	.156**	.630**	.161**	.386**	.229**	.258**	.431**
		Sig. (2-tailed)	.000	.	.000	.002	.000	.001	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_IR_D3	Correlation Coefficient	.220**	.258**	1.000	.219**	.720**	.112*	.197**	.571**	.270**	.477**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.025	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_IR_D4	Correlation Coefficient	.307**	.156**	.219**	1.000	.599**	.078	.123*	.177**	.525**	.424**
		Sig. (2-tailed)	.000	.002	.000	.	.000	.120	.014	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_IR_TO	Correlation Coefficient	.556**	.630**	.720**	.599**	1.000	.288**	.363**	.488**	.495**	.696**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_D1	Correlation Coefficient	.507**	.161**	.112*	.078	.288**	1.000	.119*	.149**	.074	.404**
		Sig. (2-tailed)	.000	.001	.025	.120	.000	.	.017	.003	.141	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_D2	Correlation Coefficient	.238**	.386**	.197**	.123*	.363**	.119*	1.000	.193**	.151**	.602**
		Sig. (2-tailed)	.000	.000	.000	.014	.000	.017	.	.000	.002	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_D3	Correlation Coefficient	.178**	.229**	.571**	.177**	.488**	.149**	.193**	1.000	.266**	.623**
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.003	.000	.	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_D4	Correlation Coefficient	.241**	.258**	.270**	.525**	.495**	.074	.151**	.266**	1.000	.679**
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.141	.002	.000	.	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_IR_TO	Correlation Coefficient	.436**	.431**	.477**	.424**	.696**	.404**	.602**	.623**	.679**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.
		N	399	399	399	399	399	399	399	399	399	399

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

			OS_DR_D1	OS_DR_D2	OS_DR_D3	OS_DR_D4	OS_DR_TO	AS_DR_D1	AS_DR_D2	AS_DR_D3	AS_DR_D4	AS_DR_TO
Spearman's rho	OS_DR_D1	Correlation Coefficient	1.000	.083	.209**	.111*	.528**	.687**	.035	.157**	.106*	.510**
		Sig. (2-tailed)	.	.098	.000	.027	.000	.000	.483	.002	.034	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_D2	Correlation Coefficient	.083	1.000	.176**	.222**	.589**	.152**	.548**	.244**	.115*	.437**
		Sig. (2-tailed)	.098	.	.000	.000	.000	.002	.000	.000	.022	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_D3	Correlation Coefficient	.209**	.176**	1.000	.316**	.567**	.226**	.187**	.560**	.271**	.390**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_D4	Correlation Coefficient	.111*	.222**	.316**	1.000	.659**	.171**	.322**	.350**	.619**	.478**
		Sig. (2-tailed)	.027	.000	.000	.	.000	.001	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	OS_DR_TO	Correlation Coefficient	.528**	.589**	.567**	.659**	1.000	.548**	.461**	.475**	.460**	.836**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_D1	Correlation Coefficient	.687**	.152**	.226**	.171**	.548**	1.000	.121*	.279**	.203**	.662**
		Sig. (2-tailed)	.000	.002	.000	.001	.000	.	.015	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_D2	Correlation Coefficient	.035	.548**	.187**	.322**	.461**	.121*	1.000	.237**	.183**	.542**
		Sig. (2-tailed)	.483	.000	.000	.000	.000	.015	.	.000	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_D3	Correlation Coefficient	.157**	.244**	.560**	.350**	.475**	.279**	.237**	1.000	.280**	.545**
		Sig. (2-tailed)	.002	.000	.000	.000	.000	.000	.000	.	.000	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_D4	Correlation Coefficient	.106*	.115*	.271**	.619**	.460**	.203**	.183**	.280**	1.000	.498**
		Sig. (2-tailed)	.034	.022	.000	.000	.000	.000	.000	.000	.	.000
		N	399	399	399	399	399	399	399	399	399	399
	AS_DR_TO	Correlation Coefficient	.510**	.437**	.390**	.478**	.836**	.662**	.542**	.545**	.498**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.
		N	399	399	399	399	399	399	399	399	399	399

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

			OS_DR_D1	OS_DR_D2	OS_DR_D3	OS_DR_D4	OS_DR_TO	AS_DR_D1	AS_DR_D2	AS_DR_D3	AS_DR_D4	AS_DR_TO
Spearman's rho	OS_DR_D1	Correlation Coefficient	1.000	.155**	.240**	.096	.428**	.440**	.147*	.179**	.100	.295**
		Sig. (2-tailed)	.	.007	.000	.100	.000	.000	.011	.002	.084	.000
		N	296	296	296	296	296	296	296	296	296	296
	OS_DR_D2	Correlation Coefficient	.155**	1.000	.178**	.154**	.590**	.256**	.345**	.240**	.098	.325**
		Sig. (2-tailed)	.007	.	.002	.008	.000	.000	.000	.000	.093	.000
		N	296	296	296	296	296	296	296	296	296	296
	OS_DR_D3	Correlation Coefficient	.240**	.178**	1.000	.350**	.677**	.258**	.233**	.533**	.323**	.511**
		Sig. (2-tailed)	.000	.002	.	.000	.000	.000	.000	.000	.000	.000
		N	296	296	296	296	296	296	296	296	296	296
	OS_DR_D4	Correlation Coefficient	.096	.154**	.350**	1.000	.716**	.193**	.326**	.367**	.602**	.568**
		Sig. (2-tailed)	.100	.008	.000	.	.000	.001	.000	.000	.000	.000
		N	296	296	296	296	296	296	296	296	296	296
	OS_DR_TO	Correlation Coefficient	.428**	.590**	.677**	.716**	1.000	.412**	.442**	.532**	.496**	.699**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.000	.000
		N	296	296	296	296	296	296	296	296	296	296
	AS_DR_D1	Correlation Coefficient	.440**	.256**	.258**	.193**	.412**	1.000	.280**	.352**	.251**	.572**
		Sig. (2-tailed)	.000	.000	.000	.001	.000	.	.000	.000	.000	.000
		N	296	296	296	296	296	296	296	296	296	296
	AS_DR_D2	Correlation Coefficient	.147*	.345**	.233**	.326**	.442**	.280**	1.000	.257**	.224**	.639**
		Sig. (2-tailed)	.011	.000	.000	.000	.000	.000	.	.000	.000	.000
		N	296	296	296	296	296	296	296	296	296	296
	AS_DR_D3	Correlation Coefficient	.179**	.240**	.533**	.367**	.532**	.352**	.257**	1.000	.312**	.744**
		Sig. (2-tailed)	.002	.000	.000	.000	.000	.000	.000	.	.000	.000
		N	296	296	296	296	296	296	296	296	296	296
	AS_DR_D4	Correlation Coefficient	.100	.098	.323**	.602**	.496**	.251**	.224**	.312**	1.000	.665**
		Sig. (2-tailed)	.084	.093	.000	.000	.000	.000	.000	.000	.	.000
		N	296	296	296	296	296	296	296	296	296	296
	AS_DR_TO	Correlation Coefficient	.295**	.325**	.511**	.568**	.699**	.572**	.639**	.744**	.665**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.
		N	296	296	296	296	296	296	296	296	296	296

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix M.

Mean ranks and sum of ranks for the Clinical and Control Groups on Visual Reproduction subtest

	Group Membership	N	Mean Rank	Sum of Ranks
Alternate Scoring System Immediate Recall				
Design 1	Clinical	17	26.65	799.50
	Control	17	34.35	1030.50
Design 2	Clinical	17	25.47	764.00
	Control	17	35.53	1066.00
Design 3	Clinical	17	22.02	660.50
	Control	17	38.98	1169.50
Design 4	Clinical	17	21.75	652.50
	Control	17	39.25	1177.50
Total Score	Clinical	17	21.98	659.50
	Control	17	39.02	1170.50
Alternate Scoring System Delayed Recall				
Design 1	Clinical	17	25.63	769.00
	Control	17	35.37	1061.00
Design 2	Clinical	17	21.55	646.50
	Control	17	39.45	1183.50
Design 3	Clinical	17	24.10	723.00
	Control	17	36.90	1107.00
Design 4	Clinical	17	24.90	747.00
	Control	17	36.10	1083.00
Total Score	Clinical	17	21.85	655.50
	Control	17	39.15	1174.50

Table continued on following page

Original Scoring System Immediate Recall				
Design 1	Clinical	17	28.00	840.00
	Control	17	33.00	990.00
Design 2	Clinical	17	27.38	821.50
	Control	17	33.62	1008.50
Design 3	Clinical	17	22.35	670.50
	Control	17	38.65	1159.50
Design 4	Clinical	17	22.18	665.50
	Control	17	38.82	1164.50
Total Score	Clinical	17	21.67	650.00
	Control	17	39.33	1180.00
Original Scoring System Delayed Recall				
Design 1	Clinical	17	25.52	765.50
	Control	17	35.48	1064.50
Design 2	Clinical	17	22.35	670.50
	Control	17	38.65	1159.50
Design 3	Clinical	17	24.83	745.00
	Control	17	36.17	1085.00
Design 4	Clinical	17	25.20	756.00
	Control	17	35.80	1074.00
Total Score	Clinical	17	22.97	689.00
	Control	17	38.03	1141.00