Development of an alternative scoring system for the Visual Reproduction subtest of the Wechsler Memory Scale -Revised

Submitted by

Therese Clark B. A. (Hons)

A thesis submitted as partial fulfillment of the requirements
for the degree of Doctor of Psychology

Psychology Department

Victoria University of Technology
St Albans, Victoria, 3021
Australia

October, 2000.

ABSTRACT

The Visual Reproduction subtest of the Wechsler Memory Scale-Revised has been widely used as a measure of memory. However, there are a number of limitations with the administration, scoring and interpretation of the subtest. A revised scoring system was developed in this study in order to address some of the deficiencies in the original scoring system. The revised scoring system was found to have very good reliability and this reliability was at least equal to that found for the original scoring system. There was also a high correlation between the two scoring systems, indicating that both systems generated a similar grading of memory performance in an elderly sample. However, further research is needed to extend the psychometric information available and to develop a normative base. An Index of non-verbal memory performance was also derived from the revised scoring system in order to provide an indicator of deficient non-verbal memory. This Index was able to discriminate between persons with left and right hemisphere lateralised lesions and therefore has the potential to provide a diagnostic indicator of non-verbal memory dysfunction. Although further research is needed in order to validate this Index in other samples, this approach provides an alternative to the development of new "non-verbal" memory tests. In order to extend the clinical information available from the Visual Reproduction Subtest, cueing, recognition, and perceptual match procedures were also developed and standardised. The results obtained with these measures indicated that a number of individuals were able to recall substantially more information than was found using free recall procedures. Further research is needed in order to obtain normative information about performance on these tasks in normal and clinical populations.

DECLARATION OF AUTHORSHIP

" I declare that this thesis does not contain any material written or published elsewhere by another person except where reference is made in the text of the thesis. I further declare that no persons work has been used without due acknowledgement in the main text of the thesis."

"This thesis has not been submitted for the award of any other degree or diploma in any other tertiary institution."

"I further declare that this study has adhered to the ethical principals as established by the Psychology Ethics Committee of Victoria University."

Therese Clark
30/9/2000

ACKNOWLEDGEMENT

I would like to thank my supervisor, Dr Peter Dowling, for his guidance and assistance through this project. My sincere thanks go to those who participated in this study and volunteered their time. I would also like to thank my family, without whose invaluable support this work would not have been possible. Finally, my greatest thanks go to my husband, Richard, for his patience and support. He too has ridden the lows and highs of thesis preparation with unconditional support.

TABLE OF CONTENTS

TABLE OF CONTENTS CONT.

TABLE OF CONTENTS CONT.

TABLE OF CONTENTS CONT.

TABLE OF CONTENTS CONT.

TABLE OF CONTENTS CONT.

LIST OF TABLES

LIST OF TABLES CONT.

LIST OF TABLES CONT.

LIST OF FIGURES

INTRODUCTION

According to Tranel and Damasio (1995) memory "refers to knowledge that is stored in the brain and to the processes of acquiring and retrieving such knowledge" (p. 31). Many current theories of memory tend to view it as comprising a number of components that are inter-related and that are mediated by a number of connected neural systems (Baddeley, 1995; Curran & Schacter, 1997; Leng & Parkin, 1990; Squire, Knowlten & Musen, 1993; Zola-Morgan & Squire, 1993).

A Theoretical Approach to Memory

A number of different models and theories of memory have been proposed to explain how memory traces are encoded, stored and retrieved (Atkinson & Shiffrin, 1968; Baddeley, 1986; Craik & Lockhart, 1972; Tulving, 1972, 1985). Detailed descriptions of these models are available elsewhere (Atkinson & Shiffrin, 1968; Baddeley, 1986, 1990; Squire, 1982, 1986, 1987; Squire & Butters, 1992; Squire & Cohen, 1984; Tulving, 1985) and only a brief overview will be presented here.

A Model of Memory Processing

Many early models of memory were based on an information-processing approach, in which information was hypothesised as passing through a number of distinct stages. A classic example of this approach was the gateway theory of memory organisation proposed by Atkinson and Shiffrin (1968). In this model, sensory information was thought to initially enter iconic or echoic memory. This information was stored for very brief periods and decayed rapidly unless it was attended too and transferred to short-term memory (STM).

Short-term memory was considered capable of storing small amounts of information for approximately 30 seconds; although STM was also thought to be extremely sensitive to interference from competing information. If the material in STM was rehearsed or encoded, it was transferred to long-term memory (LTM).

Long-term memory was seen as having an unlimited storage capacity, with an ability to hold memories permanently. It was said to be characterised by a high level of organisation, with information being available for later retrieval through short-term memory.

The concept of STM was modified by Baddeley and Hitch (1974) to account for research findings that where inconsistent with the notion of a single short-term storage system. They proposed that there were a number of temporary storage systems that were capable of storing different types of information for brief periods of time. They replaced the concept of STM with that of working memory to represent the activity of these storage systems. In this model, working memory consisted of a central executive that organised the activity of the temporary storage systems, including the phonological loop (auditory-verbal information) and the visuo-spatial scratch pad (visual-spatial information).

The information-processing approach to memory organisation also assumed that the processing of information consisted of three sequential processes, including encoding (the process by which sensory information was transformed into a memory), storage (the persistence of the memory over time) and retrieval (the process of accessing stored information). Prominent theories of information encoding, storage and retrieval included those of Craik and Lockhart (1972), Mandler (1980), and Tulving and Thompson (1973).

Organisation and Storage in Memory

Long-term memory has been typically understood as comprising multiple components. Several distinctions have been postulated based on clinical and experimental studies including: 1) declarative memory and procedural memory (Cohen & Squire, 1980), 2) episodic memory and semantic memory (Tulving, 1972; 1987, 1991), 3) explicit memory and implicit memory (Squire, 1992), and 4) verbal memory and non-verbal memory (Milner, 1958; 1971). These distinctions will be briefly described.

Episodic and semantic memory.

Tulving (1972) proposed that episodic and semantic memories were different types of knowledge systems in LTM. Episodic memories were said to be autobiographical, with memories for past events and experiences being linked with the time they occurred. According to Tulving (1972) semantic memories involved context-independent general knowledge, including the knowledge of words, symbols, concepts and rules.

Declarative and procedural memory.

Cohen and Squire (1980) proposed that memories could be categorised as either declarative or procedural in nature. Declarative memories included facts and data from previous experiences. Semantic memory and episodic memory have been considered declarative memories by some authors, because both are derived from previous experiences, both can be consciously retrieved and both can be used in many contexts (Baddeley, 1995; Cohen & Eichenbaum, 1993; Squire et al., 1993). In contrast, the term procedural memory was used to refer to skill acquisition, rule learning, habit formation and procedural learning. It was postulated that procedural memories could not be consciously retrieved and used in different contexts other than those that were present when the information was learned (Cohen & Eichenbaum, 1993).

The procedural memory concept was renamed non-declarative memory and expanded by some authors to include priming, simple classical conditioning and non-associative learning (Squire, 1987; Squire et al., 1993). Conversely, Cohen and Eichenbaum (1993) suggested that these additional aspects of memory (priming, simple classical conditioning and non-associative learning) could be seen as different components of the procedural memory system.

Implicit and explicit memory.

All tasks used to assess procedural/non-declarative memory could be considered implicit memory tasks. In an implicit memory task, an individual's performance can be influenced by previous events, but specific episodes of learning cannot be consciously identified (Graf & Schacter, 1985; Squire, 1992). The term explicit memory has been used to refer to tasks that assess declarative, episodic and semantic memories. These memories are capable of being accessed in different ways to how they were encoded and there can also be a recollection of having learned the information at some point in time.

Verbal and non-verbal memory.

A distinction has also been made between declarative memory for material that is non-verbal and for material that is verbal in nature (Milner, 1971). It has been suggested that material of a non-verbal nature is stored in a different area of the brain than material that is of a verbal nature (Squire & Butters, 1992). This distinction will be explored in more detail in a later section.

Current Understanding of the Foundations of Human Memory

The Neurobiology of Memory

For the encoding, storage and retrieval of memories to occur, information must leave a trace in mental processes. This memory trace has been postulated to result from some type of structural and/or biochemical modification.

Studies of habituation, sensitisation and classical conditioning in the marine mollusc, *Aplysia californica*, have supported the idea that the consolidation of information into a memory trace could result from alterations in the strength of synapses. Kandel and Schwartz (1982) and Shepherd (1994) provide an extended discussion of these studies. In humans, similar changes in synaptic strength have been found as the result of processes called Long Term Potentiation (LTP) and Long Term Depression (LTD). These processes could potentially provide the substrate for some forms of memory in humans because they have been shown to lead to long-term physiological changes in the human brain.

Long Term Potentiation (LTP) refers to the long-term increase in the excitatory depolarisation potential of a post-synaptic cell that can result from extended stimulation of a pre-synaptic cell (Bliss & Lomo, 1973). This increase in excitatory potential can continue for a considerable period, resulting in changes in the strength of synapses. In the human brain, LTP has been demonstrated in the CA3 and CA1 areas of the hippocampus, the dentate gyri and areas of the neocortex. All of these areas have been shown to be involved in memory (Shepherd, 1994). At a cellular level, LTP has been shown to result from a cascade of events involving glutamate, NMDA receptors, non-NMDA receptors, sodium ions, calcium ions and protein kineases (Shepherd, 1994). Acetylcholine, Serotonin and Neuropeptides such as Vasopressin have also been implicated (Kopelman, 1992).

A complementary process to LTP, called Long Term Depression (LTD), has been demonstrated in the human cerebellum. Like Long Term Potentiation, Long Term Depression (LTD) can also result in physiological changes at a neural level. LTD refers to the finding that stimulation of mossy or parallel fibres with climbing fibres results in long-lasting depression in Purkinje cell responses (Thompson, 1986). LTP and LTD could provide the neural basis for at least some kinds of human memories.

Studies have also suggested that changes in gene transcription may produce prolonged alterations in a cells excitatory potential (Shepherd, 1994). Indeed, studies have shown that learning does not occur when protein and RNA synthesis is reduced, suggesting a role for these substances in memory (McGeer, Eccles & McGeer, 1978). Further research is needed to establish if and how gene transcription is involved in human memory.

The Neuroanatomy of Memory

The gross anatomical basis of memories has been derived from animal models (Mishkin, 1978, 1982), from observations of individuals with localised brain damage (Squire, 1986; Zola, 1997), from ablation studies in animals (Damasio, Graff-Radford, Eslinger, Damasio & Kassell, 1985; Zola-Morgan, Squire & Amaral, 1989) and more recently from functional imaging studies (Squire, Amaral & Press, 1990).

Declarative Memory

Anterograde memory: the temporal lobes.

The medial temporal lobes (hippocampus, entorhinal cortex, perirhinal cortex, parahippocampal cortex) have been demonstrated to be involved in the acquisition of declarative memories (Tranel & Damasio, 1995; Squire & Zola-Morgan, 1991; Zola-Morgan & Squire, 1993). A significant anterograde amnesia has been found with damage limited to the CA1 region of the hippocampus, although concurrent damage to the entorhinal, perirhinal and para-hippocampal cortices can produce a more severe amnesia (Mishkin, 1978; Squire, 1986; Zola, 1997). Damage restricted to the perirhinal and para-hippocampal cortices has also been found to produce memory impairment (Squire, 1992), although damage to the fornix has not been shown to produce long lasting memory impairment in monkeys (Zola-Morgan et al., 1989). Thus, both human studies and primate research has suggested that the hippocampus and the surrounding cortex are the most crucial regions of the temporal lobe for memory function.

The hippocampal complex in each hemisphere have been postulated to be specialised for different types of declarative memory (Milner, 1971). In this view, the left temporal region has a primary role in verbal memory and non-verbal memory is mediated by the right temporal region (Squire & Butters, 1992). Indeed, removal of, or damage to, 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; Giovagnoli & Avanzini, 1999; Herman, Connell, Barr & Wyler, 1995a; Ivnik, Sharbrough & Laws, 1987; Rausch & Babb, 1993; Saling et al., 1993).

Damage to, or removal of, structures in the right temporal lobe has also been shown to impair the learning of and delayed recall of visual and spatial information (Giovagnoli & Avanzini, 1999; Gliebner, Helmstaedter & Elger, 1998; Glosser, Deutsch, Cole, Corwin & Saykin, 1998; Helmstaedter, Pohl, Hufnagel & Elger, 1991; Milner, 1965; Morris, Abrahams & Polkey, 1995b; Piguet, Saling, O'Shea, Berkovic & Bladin, 1994; Smith & Milner, 1989; Trenerry, Jack, Cascino, Sharbrough & Ivnik, 1996). However, not all studies have found an association between tasks used to measure non-verbal memory and the right temporal lobe (Barr et al., 1997; Dobbins, Kroll, Tulving, Knight & Gazzniga, 1998; Helmstaedter, Pohl & Elger, 1995; Ivnik et al. 1987; Lencz et al., 1992; Naugle, Chelune, Luders & Awad, 1991; Piguet et al., 1994; Rausch & Babb, 1993; Saykin, Gur, Sussman, O'Connor & Gur, 1989). The association between non-verbal memory and the right temporal lobe has been found more consistently when delayed recall has been examined or when the task has involved stimuli that are difficult to verbally encode, such as abstract designs or faces (Eadie & Shum, 1995; Jones-Gottman 1986a; 1986b; Morris, Abrahams, Baddeley & Polkey, 1995a; Warrington & James, 1967).

Anterograde memory: the diencephalon.

The thalamus (particularly the dorsomedial and anterior nuclei) plus the mammillary bodies and their connecting tracts (mammillothalamic tract and ventroamygdalofugal pathway) have been linked to episodic memory (Butters & Stuss, 1989; Kopelman, 1995a; Mayes, Meudell, Mann & Pickering, 1988; Squire, 1992; Squire & Zola-Morgan, 1988; Tranel & Damasio, 1995). Although the precise roles of these regions have not been well specified, these sites have been implicated in the neuropathology of Wernicke-Korsokoff Syndrome and other forms of diencephalic amnesia. Of note, most of these regions have connections to the hippocampus or entorhinal cortex.

Anterograde amnesia: the basal forebrain.

The basal forebrain nuclei (septal nuclei, the diagonal band of Broca, and the nucleus basalis, substansia innominata) have also been implicated in episodic memory. As this region is involved in delivering acetylcholine to the hippocampus and many areas of the cerebral cortex, damaged can result in defects in memory functioning (Damasio et al., 1985).

Anterograde amnesia: the frontal lobes.

Lesions of the frontal lobes may also result in reduced memory functioning (Squire & Butters, 1992). Preliminary studies have suggested that different regions of the frontal lobes may be involved in different aspects of memory functioning. For example, the dorsolateral frontal region may be involved in recalling how many times something has happened (frequency) and how long ago an event took place (recency) (Anderson, Damasio & Tranel, 1993; Smith & Milner, 1988). Further research is required to understand the role the frontal lobes and their connections play in different types of memory functioning.

Semantic Memory

The storage area for semantic memories appears to be relatively independent of the mesial temporal memory system (Squire, 1987; Tulving, 1985; Zola-Morgan & Squire, 1993). Research has indicated that deficits in semantic memory are associated with the anterolateral aspects of the temporal lobe, particularly in the left hemisphere (Hodges, Salmon & Butters, 1992; Snowden, Griffiths & Neary, 1994; Snowden, Neary, & Mann, 1996; Snowden, Neary, Mann, Goulding & Tsta, 1992).

Retrograde Amnesia

The hippocampal system does not seem to be fundamentally involved in the retrieval of previously learned information and so it is not implicated in retrograde amnesia (Tranel & Damasio, 1995). Damage to the CA1 field of the hippocampus alone has not been found to produce a severe retrograde amnesia (Zola, 1997). However, if structures surrounding the CA1 region are also damaged, a severe and temporally graded retrograde amnesia has been found (Zola, 1997). The non-medial temporal region (the anterior, inferior and lateral portions of the temporal lobe) appears to be important in the retrieval of past memories (Hodges, 1995; Jones & Tranel, 1993; Tranel & Damasio, 1995). Studies of individual cases have also implicated the medial diencephalon, mammillary bodies, fornix and basal forebrain structures in retrograde amnesia (Hodges, 1995).

Implicit Memory

The basal ganglia and the cerebellum are thought to be involved in the acquisition and retrieval of non-declarative/procedural memories, although knowledge of the precise role of these areas in memory functioning is still limited (Squire et al., 1993; Thompson, 1986; Tranel & Damasio, 1995). This system is relatively independent and damage to these areas appears to have no effect on the functioning of the medial temporal/declarative memory system (Cohen & Squire, 1980).

Memory Dysfunction in Neurological Disease

Memory problems are a common sequel of neurological trauma and disease. They are also seen in affective disorders. Disruption of memory function, such as occurs in amnesic disorders and dementia syndromes, can have devastating effects on independent living because of the importance of learning and memory in most aspects of life. With neurologically based memory disorders, a complete loss of all forms of memory is rare and memory for some kinds of information is preserved or less affected. Additionally, memory difficulties can occur secondary to impairment in other cognitive functions such as attention.

The Amnesic Syndrome

The amnesic syndrome represents the severe end of the spectrum of memory disorders. Although there can be a great deal of heterogeneity in individual presentation, the key elements include:

1) A marked anterograde amnesia. This is an impairment in the recall of autobiographical information from the period after the onset of brain trauma. Typically, immediate memory is preserved, but a deficit in delayed declarative memory is present. Although there are many different etiologies that can cause an amnesic syndrome, an anterograde amnesia is characteristic of almost all (O'Connor, Verfaellie & Cermack, 1995; Parkin & Leng, 1993).
2) A retrograde amnesia, that is, an inability to recall episodic information from some period of time before the onset of amnesia. Context-free information such as linguistic knowledge is usually retained. The retrograde amnesia may have a temporal gradient; older memories may be more reliably retrieved than those closer to the time of the onset of the anterograde amnesia (McCarthy & Warrington, 1990; O'Connor et al., 1995; Parkin & Leng, 1993).
3) Intact classical conditioning, motor-skill acquisition, perceptual learning, and priming is found in many amnesic individuals (O'Connor et al., 1995; Parkin & Leng, 1993).

In addition to these elements, a normal attention span, normal intelligence, normal perceptual processing and knowledge of personal identity have all been included as aspects of the amnesic syndrome. Confabulatory tendencies, personality changes, executive deficits, impaired insight, behavioral and cognitive changes have been found in some, but not all cases of the amnesic syndrome (Hodges, 1995; Kopleman, 1998; O'Connor et al., 1995; Parkin, 1984).

Subtypes of the Amnesic Syndrome

Individuals with amnesia tend to be divided into those with primary damage in the 1) medial temporal region, 2) in the diencephalon, 3) in the frontal lobe, or 4) those with a non-neurological basis, that is, psychogenic amnesic (Feher & Martin, 1992; McCarthy & Warrington, 1990; Parkin & Leng, 1993). The characteristic features associated with each etiology are described in Butters and Stuss (1989), Hodges (1995), Kopleman (1995a, 1995b, 1998), O'Connor et al. (1995) and Parkin (1984).

Although amnesic syndromes have been divided into subcategories based on their primary neuropathological locus, the issue of amnesic subtypes has been a controversial area (McCarthy & Warrington, 1990; O'Connor et al., 1995; Feher & Martin, 1992). It has been argued by some that there is only one core amnesic syndrome (e.g. Weiskrantz, 1985). In contrast, others have argued that there are significant differences among people with amnesia (e.g. Butters, 1984; Butter et al., 1988). The issue of amnesic subtypes remains unresolved.

Memory Dysfunction in other Neurological Disorders

Other neurological conditions are also associated with memory dysfunction, including all dementia syndromes, degenerative conditions such as multiple sclerosis, and acquired brain injuries. The memory loss in these cases has some features that are characteristic of the amnesic syndrome, but there are usually some additional areas of deficit in memory and other cognitive functions.

Dementias have been characterised as "cortical" (e.g., Alzheimer's Disease, Fronto-Temporal Dementia) or "subcortical" (e.g., Huntington's Disease, Parkinson's Disease, Progressive Supranuclear Palsy) depending on the primary site of neuropathological changes (Lezak, 1995). Differences in the pattern of memory disorders can be seen amongst subcortical and cortical dementia syndromes. These patterns share similarities and differences with the amnesic syndromes. A comparison of amnesic and dementia syndromes illustrates the variety of presentations of those with memory impairment.

Short-term and Long-term Declarative Memory

Anterograde amnesia or a deficit in episodic memory has been identified as a cardinal diagnostic feature of dementia syndromes and amnesic disorders. However, there can be differences in the episodic memory difficulties found in these groups. For example, many individuals with amnesia have been found to have intact immediate memory, whilst having extremely poor long-term memory (Baddeley, 1995; Scoville & Milner, 1957). In contrast, immediate memory, immediate span and working memory can often be impaired along with long-term memory in the cortical dementias, such as Alzheimer's Disease (Brandt & Rich, 1995).

The anterograde amnesia seen in some cortical dementias such as Alzheimer's Disease may be (or become over time) as pervasive as that seen in the severe amnesic syndrome. However, the episodic memory impairment in other dementia syndromes may not be as severe. For example, those with a subcortical dementia syndrome may have poor immediate and delayed free recall, but can have normal performance on recognition tasks, suggesting a retrieval difficulty as the basis of their poor free recall (Brandt & Rich, 1995; Cummings, 1990; McPherson & Cummings, 1996). In contrast, those with an amnesic disorder such as Wernicke-Korsakoff Syndrome and those with a cortical dementia such as Alzheimer's Disease, typically show little reliable benefit from cueing or recognition formats on episodic memory tasks (Bondi, Salmon & Kasniak, 1996). This suggests that an encoding or storage difficulty may be the basis of their poor performance on free recall tasks.

Retrograde amnesia is a characteristic feature of both the amnesic disorders and some dementia syndromes. This retrograde amnesia typically has a temporal gradient, with the exception of Huntington's disease and Herpes Simplex Encephalitis (where the retrograde amnesia is pervasive for all periods of life). The retrograde amnesia seen in most dementia syndromes (for e.g., Alzheimer's Disease) generally has a flatter gradient than that seen in amnesic disorders (Hodges, 1995).

Semantic and Episodic Memory

It has been proposed that episodic and semantic memories can be dissociated in individuals with amnesia (Feher & Martin, 1992). Indeed, preservation of previously acquired semantic memories has been found in most amnesics with the exception of amnesia as the result of Herpes Simplex Encephalitis (which typically results in a pervasive semantic memory deficit). However, this preservation may be limited to the semantic memories acquired before the onset of brain damage and not extend to the learning of new semantic information (Parkin & Leng, 1993). Indeed, the results of more recent research have tended to suggest that many of those with amnesia have impaired new learning for episodic *and* semantic information (Baddeley, 1995; Gabrieli, Cohen & Corkin, 1988; Squire, 1987).

Semantic memory can be affected in the dementia syndromes, but to varying degrees. Semantic memory is mildly impaired in the early stages of Alzheimer's Disease with a relatively greater impairment in episodic memory. However, in Semantic Dementia, semantic memories are severely impaired at the outset, whilst episodic memory may be relatively preserved (Snowden, Goulding & Neary, 1989).

Declarative and Procedural Memory

Individuals with severe amnesic syndromes often show normal procedural learning in contrast to their severe impairments in episodic and/or semantic memory (Cohen & Squire, 1980; Corkin, 1968, 1984; Squire, 1986). The distinction between declarative and procedural memory has also been demonstrated in demented individuals. For example, individuals with Alzheimer's disease have impaired declarative learning (although episodic memory may be more greatly affected in the early stages than semantic memory). Conversely they show intact performance on some motor and perceptual tasks assessing procedural learning (Brandt & Rich, 1995). However, in other dementia syndromes, typically subcortical (e.g. Huntington's Disease), impaired performance can be found on procedural learning tasks. Thus, there can be variability in patterns of performance on procedural learning tasks amongst individuals with different dementing syndromes.

Implicit and Explicit Memory

Tasks included under the banner of implicit memory, such as priming and skill learning, have been found to be performed normally in people with an amnesic disorder, in contrast with their impaired performance on explicit memory tasks (Squire et al., 1993). Implicit memory, particularly on motor based priming tasks, motor skill learning and perceptual processing, is also typically unimpaired in those with Alzheimer's Disease, in the context of impaired episodic memory. In contrast to those with amnesia and other forms of dementia, motor, visuo-motor and perceptual skill acquisition is typically impaired in Huntington's Disease (Brandt & Rich, 1995).

Although amnesic individuals can perform near normally on priming tasks, impaired lexical, semantic and pictorial priming can be seen with some forms of dementia (e.g. Alzheimer's Disease) but can be unimpaired in others such as Huntington's Disease (Brandt & Rich, 1995). This illustrates further the variability in memory difficulties displayed across amnesic and dementing conditions.

Thus, whilst there are similarities in memory functioning between some forms of amnesia and dementia, noteworthy differences can be observed. Memory assessment procedures need to be able to address these different aspects of memory impairment.

Assessment of Memory

Memory assessment is conducted in both clinical practice and research. Clinically, memory assessment can be conducted for a number of reasons including:

- to identify a deficit;
- to determine the nature of the deficit, for example, non-verbal or verbal, recall or recognition;
- to distinguish if a deficit is organic and/or functional in origin;
- to determine whether a deficit is due to other factors such as poor motivation, poor concentration or reduced intelligence;
- to determine a pattern or profile of intact versus impaired memory skills;
- to determine if there is any change in functioning over time;

- to predict the persons memory functioning in everyday life;
- to develop a rehabilitation program; and
- to monitor the effects of rehabilitation (Parkin & Leng, 1993; Lezak, 1995).

Memory assessment for the purposes of research often has different aims which may include investigating models of memory and ascertaining different patterns of memory disorders and how these relate to brain lesions.

Theoretically, distinctions have been made between declarative and procedural memory (Cohen & Squire, 1980). However, in clinical practice memory assessment mainly examines declarative memory, and so this area will be the primary focus of this review. Declarative memory assessment can be further refined to include tasks that assess verbal memory and those that assess non-verbal memory.

Declarative Memory

Anterograde Memory

Verbal episodic memory assessment.

Assessment of verbal memory has typically involved serial recall, free recall and paired-associate tasks. Serial recall tasks, such as the repetition of digits, letters and syllables have been used to assess immediate memory capacity. Such tasks require the participant to recall items presented in a list in the exact order in which they were presented. Digit span tasks are commonly used for this purpose.

Free recall tasks typically involve the presentation of a list of words or a story. Word learning tests that are commonly used in clinical practice include the Rey Auditory Verbal Learning Test (Rey, 1964), the California Verbal Learning Test (Delis, Kramer, Kaplan & Ober, 1987), the Selective Reminding Test (Buschke & Fuld, 1974) and the Hopkins Verbal Learning Test (Brandt, 1991). In these tasks, a list of words is presented after which the participant is required to recall as many words as possible (regardless of the order of presentation). This provides information on immediate memory capacity. Repeated exposure to a list over a number of trials assesses learning. The inclusion of a distracter list after the last recall trial enables the relative contributions of proactive and retroactive interference to be determined and delayed recall of the list provides a measure of long-term retention. Many also include a task to assess delayed recognition of the material.

Paragraph recall, such as the Logical Memory Subtest of the WMS and WMS-R (Wechsler, 1945, 1987) and the Babcock Story Recall Test (Rapaport, Gill, & Schafer, 1968) have been used to assess verbal memory. Although generally found to be very reliable, the scoring of story recall tasks has been criticised. For example, these scoring systems often do not address substitutions, omissions, additions, changes in the sequence of information and guidelines often do not indicate the speed or style of administration (Lezak, 1995; Mitrushina, Boone & D'Elia, 1999).

Saling (1998) proposed that verbal memory was not a unitary function but instead comprised several levels of cognitive processing. Thus, verbal memory tasks were not considered to be equivalent for assessing verbal memory because they were different in their composition and their reliance upon cognitive processes. Saling, O'Shea and Berkovic (1995) referred to story recall tasks as "language-rich" verbal memory tests, in that they shared a high correlation with measures of language competence. Language rich memory tasks such as these were thought to be dependent on the left neocortex, but not specifically on the hippocampal region. Indeed, deficits on story recall tasks have been consistently found in groups with left hemisphere compromise (Bornstein, Pakalnis & Drake, 1988; Herman, Wyler, Somes, Berry & Dohen, 1992). However, those with left and right hippocampal sclerosis have not been found to differ in their performance on story recall tasks (Saling, 1998).

Paired associate learning tests have also been commonly used to assess aspects of verbal memory. This type of task involves the presentation of pairs of words that the participant is required to learn. In the experimental trial, the participant is given the first word and required to recall its pair. Many of these tasks, such as the versions on the Wechsler Memory Scale and its revision (Wechsler, 1945; 1987) include "familiar" pairs, based on well-learned associations, and "novel" pairs, with no previous association.

Saling et al. (1995) suggested that the learning of related of "familiar" word pairs was a "language rich" form of verbal memory assessment. Learning unrelated or novel word pairs, was considered to represent a "language poor" form of verbal memory assessment because of the lack of association with language competence. In this view, the learning of unrelated word pairs was thought to place comparatively less demand on language functioning and was therefore associated with the functioning of the left hippocampus as compared to the left medial temporal lobe.

There does not appear to be a consensus as to whether paired associate learning tasks should include "familiar" pairs or only "novel" pairs. Saling et al. (1995) found that the recall of word pairs that were devoid of syntactic or grammatical connection, like "novel" pairs, was impaired in persons with left hippocampal sclerosis. However, pairs that did have some syntactic or semantic relationship, like "familiar" pairs, were easily learnt (Saling et al., 1995). Thus, "novel" pairs may be more effective in identifying impairments in the hippocampal memory system. This has been supported by the finding that in individuals with neuronal loss/sclerosis, a verbal memory impairment has been found more consistently on paired associate learning tasks using "novel" pairs rather than on language rich memory tasks such as prose recall (McMillan, Powell, Janota & Polkey, 1987; Rausch & Babb, 1993; Saling et al., 1993).

Paired associate tasks incorporating only unfamiliar pairs may be useful for investigating the integrity of the left hippocampal memory system. However, there would appear to be some clinical benefit in examining the difference between new learning and "refreshing" of old associations (i.e. priming), particularly in the moderately to severely impaired when the use of only hard pairs can result in little recall. This poverty of immediate recall makes testing of long term retention very difficult. Thus, the inclusion of only "novel" pairs or both "novel" and "familiar" pairs may depend on the goal of assessment.

The previously mentioned tasks have been primarily procedures involving free recall. Although paired associate learning involves a cueing procedure and word list tasks have commonly included delayed recognition procedures, very few tasks have systematically included an assessment of recognition memory. The value of the free recall/recognition comparison is the opportunity it provides to understand the contributions of encoding, storage and retrieval on a memory task. A comparison of performance on a free recall and a recognition task using the same material can be useful in order to make inferences about encoding, storage and retrieval mechanisms.

Non-verbal episodic memory assessment.

Tests of non-verbal memory have typically involved the serial recall of visual-spatial information, recognition of visual material, design recall and design reproduction.

Immediate/working memory tasks involving non-verbal stimuli have typically entailed the presentation of a sequence of dots or blocks that have to be reproduced immediately after their presentation (Milner, 1971) or asked for the individual to report the number of dots presented in an array (Kimura, 1963). These tasks have been generally found to be effective in measuring immediate memory spans.

Tasks comprising design reproduction and recall have been the most common form of non-verbal memory tasks. Such tasks would include the Visual Reproduction subtest of the WMS and WMS-R (Wechsler, 1945, 1987), the Rey Osterrieth Complex Figure Test (Rey, 1941), the Rey Complex Figure Test (Meyers & Meyers, 1995), the Biber Figure Learning Test (Glosser, Goodglass & Biber, 1989), the Rey Visual Design Learning Test (Spreen & Strauss, 1991), the Visual Spatial Learning Test (Malec, Ivnik and Hinkeldey, 1991), The Brief Visuo-Spatial Memory Test (Benedict & Groniger, 1995), and the Benton Visual Retention Test (Benton, 1992). All these tests involve a visuo-motor response and require adequate visual perceptual skills. Thus, scoring and interpretation could be confounded by impairments in construction, reduced motor skill or visual perceptual disturbance. This could potentially complicate the interpretation of defective performance on these tasks (Heilbronner, 1992).

The inclusion of a recognition task may help to overcome the potentially confounding effects of poor motor skills or impairments in construction because they place fewer demands on a motor response. However, only the Biber Figure Learning Test (Glosser et al., 1989), the Rey Complex Figure Test (Meyers & Meyers, 1995), and the Rey Visual Design Learning Test (Spreen & Strauss, 1991) have included a recognition task as part of standard administrative procedure. The recognition trial of the Biber Figure Learning Test and the Rey Visual Design Learning Test requires some type of visuo-motor response, either reproducing a design (Biber Figure Learning Test) or placing material in a visuo-spatial location (Rey Visual Design Learning Test). Perceptual deficits may also compound memory problems, which are not routinely controlled for in these two tests. In the recognition trial of the Rey

Complex Figure Test, the original stimulus is not presented in its entirety; rather, individual components of the design are presented. This may lead to some confusion for the participant, as the conditions of the original presentation are different than those of the recognition trial. It also confounds cueing with recognition. Moreover, the material used in all three tasks may be amenable to verbal encoding, thus making the interpretation of defective performance (and sound performance) problematic.

Tasks involving the recognition of visual material, rather than requiring a visuo-motor response, include the Continuous Recognition Memory Test (Hannay, Levin & Grossman, 1979), the Continuous Visual Memory Test (Trahan & Larrabee, 1988), the Figural Memory subtest of the WMS-R (Wechsler, 1987) and the Faces subtest of the Warrington Recognition Memory Test (WRMT, Warrington, 1984). The Continuous Recognition Memory Test uses easily verbalisable stimuli and therefore may be confounded by verbal encoding of the material. However, the Continuous Visual Memory Test employs ambiguous designs and the Warrington Recognition Memory Test uses faces to restrict verbal labeling. The advantage of these tests is that they can be used for evaluating non-verbal memory when free recall is impaired or when the individual cannot reproduce a design. However, they do not include free recall and recognition trials using the same material. Therefore, the relative contributions of encoding, storage and retrieval to poor performance on these tasks cannot be established.

Semantic memory assessment.

The assessment of semantic memory has typically involved tests of word meanings and word knowledge. In this respect, assessment is usually directed at retrograde memory (Lezak, 1995).

Recognition memory.

Recognition can involve the act of recognising something or it can refer to a method of testing. In a recognition memory test, the test stimuli is presented among a number of alternatives and the individual is asked to identify the test stimuli (Murdock, 1982). Traditionally, memory assessment has not incorporated recognition memory tasks. Indeed, recognition memory has been all but ignored in the clinical literature until recently (Warrington, 1984; Wechsler, 1997). The importance of this type of task is the opportunity it provides in understanding the basis of defective performance. For example, a failure to retrieve information during free recall could imply that information was not encoded, was forgotten or was stored but could not be accessed. By comparing performance on free recall and recognition tasks using the same material, information can be gained to assist in determining the basis of impaired performance. If performance is superior on a recognition task, as compared to free recall, a retrieval deficit may be postulated. If performance is equally poor on both tasks, a deficit in encoding or storage may be inferred. Tasks that can help distinguish between difficulties with encoding information or a difficulty with retrieval of information from memory are likely to be diagnostically useful because certain disorders are associated with prominent encoding or prominent retrieval deficits, (Kapur, 1987; Lezak, 1995).

Individual tests of recognition memory do exist, for example, the Warrington Recognition Memory Test (Warrington, 1984). This test provides a recognition memory task for both words (verbal memory) and faces (non-verbal memory). However, many of these tests do not evaluate both recall and recognition memory using the same material.

Retrograde Memory

The clinical approach to investigating retrograde memory has been to explore people's abilities to recall salient items from their own autobiographical (episodic) or public (semantic) history (Mayes, 1995). However, performance on these tasks has often been difficult to interpret because it has been difficult to equate the salience of memories and it is hard to establish if the person knew the public information premorbidly (Feher & Martin, 1992).

Implicit Memory

Assessment of implicit memory has involved many experimental paradigms, but few of these tasks have been used in the clinical setting (Feher & Martin, 1992; Kapur, 1988). This has been due in part to the relatively recent nature of the research involving implicit memory and the lack of development of standardised tests to assess implicit memory. Tasks used in the research literature include Word-Stem completion priming, mirror drawing, motor pursuit, reading mirror reversed words and prototype learning (Feher & Martin, 1992; Squire & Shimamura, 1997).

Issues in the Clinical Assessment of Memory

As noted previously, memory impairments are a characteristic of many neurological disorders that can have a variety of presentations. Hence, memory can not be considered a unitary ability. The assessment of memory in normal participants and in neurological conditions requires the evaluation of a number of different elements. This may involve the use of a number of independently developed memory tests or a specific memory "battery". However, for these tests to be useful, certain practical and technical criteria need to be met.

A primary concern in choosing a clinical memory test is that the content of the test adequately reflects the current theoretical understanding of memory processes. Thus, a memory task needs to include measures of immediate recall, delayed recall and recognition. Other important considerations are measures of rates of forgetting and sensitivity to interference (Mayes, 1986). Tests should also be chosen to reflect common clinical dissociations, for example, non-verbal and verbal memory. However, in the case of non-verbal memory tests, procedures need to be incorporated to control for visuo-constructive, visuo-practic and visuo-spatial components.

The statistical and normative data available are also very important in the evaluation of a test. Quantitative measures such as percentiles or standardised scores are needed for individual tests and for subtests of a memory battery so that performance can be characterised in reference to a normative group and so that performance across tasks can be compared. Scoring systems also need to be developed that minimise floor effects in clinical populations and ceiling effects in non-clinical populations.

Psychometric considerations also need to be taken into account when evaluating and choosing a test instrument. A particular problem with memory assessment is that, for the majority of clinical tests, normative information is not available for clinical or elderly groups (Loring, Lee & Meador, 1989c). A test needs to have norms derived from an adequate standardisation sample and normative data available for various clinical groups. Information needs to be available about the test's reliability (the degree of stability, consistency, predictability and accuracy in the scores, i.e., small errors of measurement) and validity (content, criterion, construct, clinical). Research and clinical evidence that the test discriminates between clinical and diagnostic groups of relevance and that it reflects differences in the severity of impairment is also needed when evaluating the usefulness of a memory measure.

A number of practical issues also need to be addressed when choosing a memory test for use in clinical practice. These include the ease of administration of the instrument and the length of the test. Clinically, ecological validity is also important. The availability of alternative forms is important if serial assessment is to be conducted. Memory measures are often administered to the impaired and the elderly and therefore the measure chosen should not place excessive demands on sensory or motor systems that can be impaired in these groups.

As well as the development of specialised tests, memory test development has involved the development of omnibus batteries to provide coverage of the many different types of memory tasks. These batteries include the Memory Assessment Scale (Williams, 1991), the Denman Neuropsychology Memory Scale (Denman, 1984) and the Rivermead Behavioural Memory Test (Wilson, Cockburn & Baddeley, 1985). Reviews of these tests are available elsewhere (Lezak, 1995; Spreen & Strauss, 1998). The Wechsler Memory Scale (Wechsler, 1945; Wechsler, 1987; Wechsler, 1997) has been the most widely used memory battery (Mitrushina et al., 1999). The scale has undergone two revisions, the Wechsler Memory Scale- Revised (WMS-R, Wechsler, 1987) and the most recent revision the Wechsler Memory Scale- Third Edition (WMS-111, Wechsler, 1997). The clinician needs to decide whether to use an omnibus battery, a number of individual tests or a combination of both.

Assessment of Memory with the Wechsler Memory Scales

Structure

The Wechsler Memory Scale (WMS)

The original Wechsler Memory Scale was published in 1945 and was one of the first attempts at clinical memory assessment. It consisted of seven subtests

(Personal and Current Information, Orientation, Mental Control, Logical Memory, Memory Span, Visual Reproduction, Associate Learning) with two alternative forms. Details of the content and scoring of these subtests is found in Table 1. The scores for the subtests on the WMS were added together to yield a global memory quotient (MQ). This process will be described in more detail in a later section.

The WMS was widely criticised on practical, theoretical and psychometric levels (Erickson & Scott, 1977; Prigatano, 1978). On a practical level, it was suggested that the WMS manual was inadequate and uninformative and that the scoring rules for Logical Memory and Visual Reproduction were too brief and imprecise, resulting in poor inter-scorer agreement (Loring & Papanicolaou, 1987).

At a theoretical level, the WMS was criticised due to the assumption that the memory quotient (MQ) adequately measured memory. The use of the single summarising measure was seen as implying that memory was a unitary and an additive phenomenon, which was not consistent with clinical and research findings. The use of a single summarising measure was also criticised because it had the potential of masking variability among subtests, which in turn could limit the clinical information that could be obtained from the test (Erickson & Scott, 1977; Herman, 1988; Loring & Papanicolaou, 1987).

The inclusion of orientation and personal information questions in the MQ of the WMS was also considered questionable, as these questions did not appear to reflect memory recall (Erickson & Scott, 1977; Prigatano, 1978). In addition, the lack of coverage of different types of memory functioning, for example, delayed memory and recognition memory and the greater number of tasks measuring verbal as compared to non-verbal memory, were identified as major weaknesses (Erickson & Scott, 1977; Herman, 1988; Loring & Papanicolaou, 1987).

Psychometrically, the WMS was criticised for the small size of the original normative sample, the lack of scaled scores for individual subtests and the lack of information about reliability. The high correlation between MQ and Wechsler IQ scores was also criticised because it raised questions about the basis of the MQ, that is, whether it measured memory or intellectual functioning. The composition of the MQ was also questioned because only three of the measures were thought to reflect memory as compared to other cognitive functions such as attention and concentration (Erickson & Scott, 1977; Loring & Papanicolaou, 1987; Prigatano, 1978).

Table 1

Subtest Content of the Wechsler Memory Scale

Subtest Name	**Content**	**Scoring**
Personal and Current Information	Six general and personal information questions.	A point for each correct answer.
Orientation	Five questions relating to time and place.	A point for each correct answer.
Mental Control	Three items to assess attention including counting backwards from 20, recitation of the alphabet and counting by threes.	A point if there was one error or less within the specified limit. Bonus points were awarded for quick performance.
Logical Memory	Immediate recall of two orally presented prose passages.	Credit for verbatim recall of 24 story units for Story 1 and 22 story units For story 2. Final score was an average of recall on the two stories.
Memory Span	Serial recall of digits, either in the same order (Digits Forward) or the reverse order (Digits Backward).	Number of digits repeated correctly. Maximum score of 8 on Digits Forward and 7 on Digits Backward.
Visual Reproduction	Immediate recall of geometric designs that were shown for 10 seconds. Cards A and B contained one design, and card C had two drawings.	Three points could be awarded for Card A, 5 points for Card B and 14 points for Card C.
Associate learning	Recall of 10 word pairs. Six of these pairs were related and four were novel. Word pairs were presented three times with recall tested immediately after each trial.	Score was the number of correct responses on the easy pairs divided by two, plus the total number of correct responses on the hard pairs.

In order to remedy some of the limitations of the WMS, two variations were developed. Russell (1975) produced a revision that incorporated delayed recall trials for the Logical Memory and Visual Reproduction subtests. This revision continues to be widely used and continues to have normative data published for it (Russell, 1988). However, Russell also included cues for both the subtests and no normative data was provided without cueing, making it difficult to compare this variation with the original version.

The Boston Revision of the WMS (Milberg, Hebben & Kaplan, 1986) included all the elements of the original, but added delayed recall and immediate and delayed recognition trials for the Visual Reproduction, Logical Memory, and Paired Associates subtests. The Visual Reproduction subtest also incorporated a copy trial following immediate recognition and a perceptual match trial following delayed recognition. Although these modifications addressed many of the problems with the content of the original, no detailed clinical data was reported. The inclusion of immediate recognition trials and an immediate copy trial for the Visual Reproduction Subtest also allowed additional exposure to the designs. Thus, performance on delayed recall with the Boston Revision was not directly comparable to the WMS version of the subtest and in fact reflected multiple exposure to the material, confounding interpretation of delayed recall.

The Wechsler Memory Scale-Revised (WMS-R)

The first revision of the WMS, the Wechsler Memory Scale-Revised, was published in 1987. Major changes to the original scale included the extension of the age range, an increase in the size of the normative sample, the replacement of the MQ with five composite scores, the addition of three new visual memory subtests, the inclusion of delayed recall trials for most of the subtests and the revision of scoring procedures for several subtests.

The WMS-R retained some of the WMS subtests with item and scoring changes and included three new subtests. In total the WMS-R contained nine subtests, Information and Orientation, and eight tasks of verbal and non-verbal learning and retention (Mental Control, Digit Span, Logical Memory, Verbal Paired Associates, Visual Paired Associates, Visual Reproduction, Figural Memory, Spatial Span).

A brief overview of the major changes in the WMS-R follows and the reader is referred to the manual for more information (Wechsler, 1987). Major changes to the scale included:

1) The Information and Orientation subtests were combined and no longer contributed to the general memory index;
2) The content of all other subtests was retained with some modification. The Logical Memory subtest included one of the original stories with minor modification (Story A) and a different passage, Story B. The Visual Reproduction subtest retained the first two cards from the WMS subtest and two new cards were added, one with one design and the other with two designs. The Verbal Paired Associates subtest contained eight of the original word pairs, and the two word pairs that were highly overlearned were eliminated.
3) Updated and more comprehensive scoring guidelines were included for the Logical Memory and Visual Reproduction subtests to address previous criticisms.
4) Three new subtests were added to the WMS-R, as potential measures of non-verbal memory. The Visual Paired Associates subtest required the recall of the association between colours and abstract line drawing pairs. The Figural Memory subtest involved multiple-choice recognition of shaded geometric designs. The Visual Memory Span subtest involved mimicking a sequence of taps on an array of coloured squares.
5) Delayed recall trials were added for Logical Memory, Visual Reproduction, Verbal Paired Associates and Visual Paired Associates to measure retention of information over time.

Although several of these changes were considered to be worthwhile improvements, a number of criticisms have been directed at the WMS-R (Elwood, 1991a; Loring, Lee, Martin & Meador, 1989b; Wong & Gilpin, 1993; Woodard, 1993; Zeilinski, 1993).

From a practical and clinical viewpoint, the WMS-R took longer to administer than the WMS and there was no alternative forms for serial administration (Ivison, 1993). Theoretically, the test did not encompass the range of tasks thought to encompass memory performance (which the manual acknowledged). For example, it did not assess every modality, everyday memory problems, autobiographical memory or procedural memory (Zeilinski, 1993). There was also limited recognition testing procedures for verbal and some nonverbal tasks, which prevented a direct comparison between recall and recognition testing (Loring, 1989a; Reid & Kelly, 1993; Troster et al., 1993). Although the Figural Memory subtest assessed non-verbal recognition memory and the Visual Reproduction subtest assessed non-verbal memory via a free recall procedure, no direct comparison could be made between free recall and recognition of the same material.

Psychometrically, the test was criticised because raw scores were used to generate index scores, which resulted in the subtests contributing differentially to the total score. Norms for separate subtests were also limited to percentiles for Logical Memory, Visual Reproduction, Digit Span and Spatial Span. The use of interpolated scores for some age groups and the small size of the standardisation sample were also criticised (Elwood, 1991a). These criticisms will be reviewed in more detail later.

Wechsler Memory Scale- Third Edition (WMS-111)

The Wechsler Memory Scale-Third Edition (WMS-111) was the second revision of the original memory scale and it involved major changes in the content and administration. Many of these changes were significant improvements on the earlier editions and as a result the WMS-111 has the potential to provide a more detailed evaluation of memory.

The WMS-111 contained six primary and five optional subtests (Table 2). Seven of the original subtests were retained, but the Figural Memory and Visual Paired Associates subtest of the WMS-R were eliminated. This was consistent with clinical practice and with research data that questioned the reliability of these two subtests as measures of non-verbal memory (Loring, 1989a; Wong & Gilpin, 1993; Zeilinski, 1993).

Table 2

Primary and Optional Subtests on the WMS-111

Primary Subtests
Logical Memory
Family Pictures
Verbal Paired Associates
Faces
Spatial Span
Letter-Number Sequencing
Optional Subtests
Information and Orientation
Mental Control
Digit Span
Visual Reproduction

Four new memory tasks were added in the WMS-111, namely, Family Pictures, Faces, Word Lists and Letter Number Sequencing. There were also major changes to most of the other subtests retained from the previous edition (Information and Orientation, Digit Span, Mental Control, Verbal Paired Associates, Logical Memory, Visual Reproduction). The WMS-111 not only included measures of immediate and delayed memory, but also incorporated measures of recognition for most of the memory tasks.

In briefly examining the content of the WMS-111, the Information and Orientation subtest was retained in the second revision with only one item being slightly modified. Mental Control was also retained as an optional subtest, with modification to include more items and the addition of time bonus points. The Digit Span subtest was not fundamentally altered, but shorter and longer digit sequences were added.

The Logical Memory subtest was retained as a primary subtest. There were slight wording changes in Story A, but Story B was totally replaced. Moreover, Story B was administered twice in the immediate condition, with recall being reassessed after each administration. The delayed recall trial was retained and a delayed recognition task was added. Scoring criteria were included for accuracy and thematic content, the later measuring the ability to recall major themes of the story.

The Verbal Paired Associates procedure was also retained as a primary subtest. However, all the eight word pairs from the WMS-R were replaced with new unrelated pairs, which were said to contain high imagery words. Four trials were given in the immediate condition and a recognition trial was added after delayed recall.

The Visual Reproduction subtest was retained but designated an optional subtest. Design B from the WMS-R was discarded and two new designs were added to extend the floor and the ceiling of the test. In addition to immediate and delayed recall trials, a recognition task, a perceptual discrimination task and a copy trial were added. The scoring system was significantly revised, with partial credit being awarded on almost all the scoring items.

The Faces subtest on the WMS-111 was included as a primary subtest and evaluated immediate and delayed memory for faces via a recognition procedure. Participants were required to remember 24 target faces (presented individually for two seconds) and subsequently identify them from a series of targets and distracters. After a delay, the individual was asked to identify the faces from the original set from a series of distracters.

The Family Pictures subtest involved the presentation of four different scenes of a family group (six people and their dog) undertaking an activity. Each of the scenes was sequentially presented for 10 seconds each. After all four scenes were presented, the individual was asked questions about the content (characters, spatial location of the characters and the activity being undertaken by the character) of these pictures. After a delay period the individual was again asked questions about the content of these pictures.

The Word Lists subtest involved the presentation of a list of 12 semantically unrelated words for immediate recall. This process was repeated over four trials, after which a new 12-word list was presented for immediate recall. After this interference task, and after a delay period, the individual was again asked to recall the words from the first list. The delayed trial was followed by a recognition procedure where the individual was presented with 24 words and asked to indicate if a word was on the first list.

The Spatial Span subtest assessed an individual's ability to reproduce a sequence of visual-spatial locations either in the same order or in the reverse order. It was essentially the same task as the Visual Memory span subtest of the WMS-R, although the series of spatial patterns were presented on a three dimensional board rather than on a two dimensional card.

Letter Number Sequencing involved the presentation of a string of letters and numbers of gradually increasing length (from two to eight elements). The individual was required to repeat the letter-number sequence by first repeating the numbers together in ascending order and then the letters together in alphabetical order.

Although the WMS-111 contained many different tasks to measure many different aspects of memory, the administration of all the primary subtests is likely to be time consuming. The manual suggested the core battery could be administered in 30 to 40 minutes, however, the elderly and other clinical groups are likely to take longer than this to complete the test. The constant demands on memory when administering even the core battery alone could also result in a significant risk of fatigue for the elderly and clinical groups. This may complicate interpretation of memory performance across the subtests.

Scoring and Index Composition of the Wechsler Memory Scales

Wechsler Memory Scale (WMS)

On the WMS, the scores for the seven subtests were added together to yield a raw score. After a score correction was added to this raw score according to the participants age, the corrected memory score was converted to the memory quotient (MQ) according to a table of MQ equivalents. The MQ was extensively criticised on many grounds, such as the assumption that memory could be reflected in a single score and because only three of the subtests measured memory functions that could be affected by amnesic syndromes (Erikson & Scott, 1977; Loring & Papanicolaou, 1987; Priganto, 1978).

Wechsler Memory Scale- Revised (WMS-R)

On the WMS-R, the eight subtests contributed to five indexes, which replaced the single mental quotient on the WMS. The Indexes were equivalent statistically to IQ scores, with each Index having a mean of 100 and a standard deviation of 15. To obtain these Index scores, raw scores for the individual subtests were weighted, and the sums of these weighted scores were added to yield a weighted raw score. This was then converted into an Index score according to the participant's age. The subtests corresponding to the summary indexes and their weighting are shown in Table 3.

Table 3
Summary Indexes on the WMS-R

Summary Indexes	**Subtest Contribution**
Attention/Concentration	Mental Control x 1 Digit Span x 2 Visual Memory Span x 2
Verbal Memory	Logical Memory Immediate Recall x 2 Verbal Paired Associates Immediate Recall x 1
Visual Memory	Figural Memory x1 Visual Paired Associates Immediate Recall x 1 Visual Reproduction Immediate Recall x 1
General Memory	The sum of the Verbal and Visual Memory Index weighted scores.
Delayed Recall	Logical Memory Delayed x 1 Visual Paired Associates Delayed x 2 Verbal Paired Associates Delayed x 2 Visual Reproduction Delayed x 1

According to Herman (1988), the five Indexes were generated based on factor analysis results of the original and revised scales, clinical evidence and rational considerations. However, Wechsler's (1987) original factor analysis of the scale and subsequent factor analysis of the scale did not appear to support a five-factor solution in many clinical and non-clinical groups (Elwood, 1991b; Roid, Prifitera & Ledbetter, 1988; Roth, Conboy, Reeder & Boll, 1990; Woodard, 1993).

The differential weighting of the individual subtests when deriving Index scores presents a potential problem because it could lead to an Index score representing performance on only one subtest. For example, on the Visual Memory Index, Visual Reproduction contributes up to 41 points, but Visual Paired Associates has a maximum score of 18 points and Figural Memory just 10 points. Thus, a poor performance on Visual Reproduction could dramatically reduce the Visual Memory Index score, whilst poor performance on Figural Memory may have a modest impact.

The composition of the WMS-R Indexes has also been criticised. For example, there has been some doubt as to whether the Visual Memory index adequately assesses visual memory. In fact, Loring et al. (1989b) found that the Visual and Verbal Memory Indexes were not able to predict the laterality of brain lesions accurately.

There has also been criticism of the types of scores that could be obtained for individual subtests (Elwood, 1991a). Although means and standard deviations were provided, their use was questionable for some subtests due to restrictions in the range of possible scores and low reliabilities. Percentile scores could be obtained for Visual Reproduction, Logical Memory, Visual Memory Span and Digit Span, however these were relatively crude measurements based on small samples. The absence of scaled scores also made it difficult to compare performance across subtests as each used a different scale. In addition, scores below 50 could not be awarded, and therefore the test was not helpful in discriminating individuals with severe memory impairment (Squire & Shimamura, 1987).

Wechsler Memory Scale- Third Edition (WMS-111)

There was a fundamental shift in the scoring of the new WMS-111 and in the Index structure. On the WMS-111, all raw scores for the primary subtests (and many of the supplemental subtests) were converted to scaled scores for each age group, with each having a mean of 10 and standard deviation of 3. This meant that each subtest could be given and interpreted separately and allowed a more reliable comparison across tasks. The potential bias in adding raw scores from different tasks (as is the case on the WMS-R) was also avoided by using age adjusted scaled scores.

Scaled scores for the primary subtests were added to construct eight Index scores. Table 4 outlines these new indexes and the subtests that contribute to them. Each Index has a mean of 100 and standard deviation of 15 and each can be compared with the Index scores from the WAIS-111.

Along with Index scores, percentiles and confidence intervals could be calculated for many of the WMS-111 subtests. There were also tables to analyse the discrepancies required for statistical significance between index scores, between subtest scaled scores, and the difference between index scores and the WAIS-111. These revisions add to the psychometric sophistication and the clinical usefulness of the scale. Indeed, the scoring of the WMS-111 is considerably superior to the previous editions.

Table 4

Index Composition on the WMS-111

Primary Index	**Composition**
Auditory Immediate	Logical Memory Immediate Recall Verbal Paired Associates Immediate Recall
Visual Immediate	Faces Immediate Recall Family Pictures Immediate Recall
Auditory Delayed	Logical Memory Delayed Recall Verbal Paired Associates Delayed Recall
Visual Delayed	Faces Delayed Recall Family Pictures Delayed Recall
Auditory Recognition Delayed	Logical Memory Delayed Recognition Score Verbal Paired Associates Delayed Recognition Score
General Memory	Logical Memory Delayed Recall and Delayed Recognition Score Verbal Paired Associates Delayed Recall and Delayed Recognition Score Faces Delayed Recall Family Pictures Delayed Recall
Working Memory	Spatial Span Letter Number Sequencing

It is important to note that the index scores on the WMS-111 and WMS-R were derived differently and are therefore not directly comparable. This criticism also applies to the transition from the WMS to the WMS-R. For example, on the WMS-R and WMS the General Memory Index or Memory Quotient was based entirely on immediate recall. On the WMS-111, the General Memory Index was based on scores from delayed memory and recognition measures as compared to immediate recall. The composition of the General Memory Index on the WMS-111 is consistent with current notions that retention of information over time is an indicator of everyday memory functioning and is also consistent with clinical approaches to what constitutes functional memory skills. However, it does mean that scores on this index are not directly comparable across the WMS editions.

Other potential problems can be identified in the index structure. Recognition measures from the Verbal Paired Associates and Logical Memory subtests were included as components of the General Memory Index. Delayed recall of faces, another recognition measure, was also a component of the General Memory Index. Given that free recall can be impaired, but recognition can be relatively preserved in some clinical syndromes (Lezak, 1995), combining recall and recognition in the same Index may be misleading in quantifying the level of memory dysfunction in these cases. Given the high correlation between the immediate and delayed verbal subtests, it may also be that separate indexes were unnecessary. In addition, the poor correlation found between the "visually" presented subtests would cast doubt on the utility of adding these score together to make the Visual Immediate and Visual Delayed Indices.

As well as the primary indexes, there were four experimental process indexes (Retention, Retrieval, Learning Slope and Single Trial Learning). All these Indexes were derived from the verbal memory measures. These Indexes were included to evaluate aspects of memory that were meaningful in a clinical setting. As new additions to the scale, more research needs to be conducted in order to evaluate their potential clinical usefulness.

The many psychometric changes, including the provision of scores and indexes addressed many of the criticisms made about the previous editions of the WMS.

Norms

Wechsler Memory Scale (WMS)

The original standardisation of the WMS provided data for 50 persons aged 20-29 and 46 persons aged 40-49 (Wechsler, 1945). The age range of this sample was quite limited and the sample was not adequately described. Subsequent attempts at providing norms were published for both the full WMS, adaptations of the WMS and individual subtests across normal and clinical groups. Mitrushina et al. (1999) and D'Elia, Satz and Schretlen (1989) provided a comprehensive critique of the normative studies on the WMS. Mitrushina et al. (1999) concluded that "idiosyncratic administration and scoring procedures generally limit the usefulness of the WMS normative studies" (pg. 320).

Wechsler Memory Scale- Revised (WMS-R)

Normative data on the WMS-R was derived from a sample of 316 adolescents and adults ranging in age from 16-74, matching the WAIS-R age range. Although this represented an extension of the age range from the WMS, the norms for the ages 18-19, 25-34 and 70-74 were interpolated. By limiting the age range to 74, normative data was restricted for populations who were at risk for memory impairment arising from late onset neurological disease such as dementia.

The WMS-R normative sample was stratified according to age, sex, race and geographic region with three education levels (0-11 years, 12 years and 13 years or more). However, no normative data was available by education level, which was found to be significantly related to test performance in the standardisation sample (Mitrushina et al., 1999; Zielinski, 1993).

Major criticisms of the normative data have included the small size of the normative sample and the use of interpolated norms (Butters et al., 1988; D'Elia et al., 1989; Elwood, 1991a; Loring, 1989a). Along with small sample size, there were also large sampling errors and large standard errors of measurement. According to Elwood (1991a, p 196) "As a result, the WMS-R provides only an approximate estimate of overall memory functioning."

Further normative studies have been conducted using the WMS-R. These studies were reviewed by Mitrushina et al. (1999) who recommended that the WMS-R norms were used for American individuals of average intelligence in the age ranges actually tested in the standardisation. They strongly cautioned against the use of the norms provided in the interpolated age groups of the WMS-R standardisation sample.

Normative data has been published for an Australian population in the age group 18 to 34 years (Carstairs & Shores, 2000; Shores & Carstairs, 2000). There is therefore an advantage in using the WMS-R in these age groups in an Australian population.

Wechsler Memory Scale- Third Edition (WMS-111)

The WMS-111 norms were based on a sample of 1250 health adults aged between 16-89 years across 13 age groups. This sample was half the stratified (age, sex, race, ethnicity, educational level and geographical region) random sample of 2450 adults used in the WAIS-111 standardisation. This standardisation sample was an improvement over the WMS-R, in terms of size and age range.

Despite the strength of the normative data, the WMS-111 did not provide normative data by education, despite the fact that education is well known to be positively correlated with performance on memory tests (Lezak, 1995). Additionally, when the WMS-111 was given to the standardisation sample, the optional subtests and additional verbal and non-verbal memory tasks were given during the delay period. It is possible that the norms for the delayed recall subtest in the manual may overestimate a participants memory functioning if the delay period is not filled with the same material as the standardisation sample (Senior & Douglas, 1999).

Despite these problems, the standardisation procedure, the age range included and the wealth of supportive literature represented a considerable advance from the previous editions.

Reliability

Wechsler Memory Scale (WMS)

A primary criticism of the WMS was the lack of information provided concerning reliability (Priganto, 1978). Wechsler (1945) did not report alternative form or test-retest reliability estimates or measures of internal consistency of subtests.

Wechsler Memory Scale- Revised (WMS-R)

The reliability of the Indexes and subtests on the WMS-R was generally unsatisfactory (Elwood 1991a; Kaufman, 1990). Only two of the Indexes on the WMS-R, General Memory (.81) and Attention/Concentration (.90), had a modest degree of reliability. Many of the individual subtests also did not meet liberal reliability standards. Digit Span (.88), Visual Memory Span (.81) and Logical Memory (Immediate, .74 and Delayed, .75) were the only subtests that had modest reliability. High standard errors, low reliabilities and consequently large standard errors of measurement thus limited the accuracy of scores on the WMS-R. These problems limited the WMS-R's sensitivity and any comparisons of differences between individual subtests and the Index scores were likely to be questionable.

Wechsler Memory Scale- Third Edition (WMS-111)

Reliability coefficients for the WMS-111 primary subtests and indexes were higher than for WMS-R. The average reliability coefficients for the Primary Indexes ranged from .74 (Auditory Recognition Delayed) to .93 (Auditory Immediate) with a median reliability of .87. All but Auditory Recognition Delayed (.74), Visual Immediate (.75) and Visual Delayed (.76) had test-retest reliabilities above .8.

The reliability of individual subtests was also better than those found for the WMS-R. Split half coefficients for individual subtests generally ranged between .8 and .9 with the lowest being Faces (.74). Reliability coefficients for subtest scores across all age groups ranged from .74 to .93 with a median reliability of .81. Thus, the reliability of Index and Subtest scores on the WMS-111 was a clear improvement on the previous editions.

Validity

Construct Validity- Factor Analytic Studies

Wechsler Memory Scale.

Factor analytic studies that investigated the construct validity of the WMS generally found three factor solutions including a) a general memory/learning factor composed of Logical Memory, Visual Reproduction and Paired Associate Learning, b) an attention/concentration factor composed of Mental Control and Digit Span; and an c) an Information/Orientation factor composed of Personal and Current information and the Orientation Subtests (Skilbeck & Woods, 1980). These factors were found in non-organic, organic and psychiatric populations (Priganto, 1978). These results suggested that the combination of the subtests on the WMS into a single mental quotient was inappropriate. Interpretation of the MQ was therefore difficult because it encompassed a number of different skills.

Wechsler Memory Scale- Revised.

A number of factor analytic studies using the WMS-R have been reported in the literature. However, there have been considerable discrepancies among the reported factor structures. Indeed, factor analysis studies of the WMS-R alone or with other tests in both normal and clinical populations have reported a variety of factor structures. These have included:

1) one factor solutions with no separation of immediate, delayed, verbal or visual memory factors (Elwood, 1991b, 1993; Smith et al., 1992b; Smith, Malec & Ivnik, 1992b);
2) two factor solutions including attention/concentration and immediate recall when only immediate subtests were used (Bornstein & Chelune, 1988; Roid et al., 1988; Smith et al., 1992a;Wechsler, 1987);

3) three factor solutions, including attention and concentration, immediate memory and delayed memory, when both immediate and delayed subtests were included in the factor analysis (Bornstein & Chelune, 1989; Bowden et al., 1997; Burton, Mittenberg & Burton, 1993; Roth et al., 1990; Woodard, 1993);
4) three factor solutions, including attention and concentration, verbal memory and non-verbal memory (Bornstein & Chelune, 1988, 1989; Jurden, Franzen, Callahan & Ledbetter, 1996; Larrabee & Curtiss, 1995);
5) four factor solutions (Larrabee & Curtis, 1995; Smith et al., 1992b)
6) five factor solutions when using the WMS-R and other memory and intellectual tests (Bowden, 1997; Bowden, Carstairs & Shores, 1999; Leonberger, Nicks, Larrabee & Goldfader, 1992; Nicks, Leonberger, Munz & Goldfader, 1992; Smith et al., 1992a; Smith, Ivnik, Malec & Tangalos, 1993).

Separate verbal and non-verbal factors have emerged in studies using younger participants (Bornstein & Chelune, 1989; Chelune & Bornstein, 1988), participants with well-defined unilateral lesions (Bowden, 1997), in a normal population (Bowden, Carstairs & Shores, 1999), in a re-examination of the WMS-R standardisation sample and in a clinical group from an addiction recovery unit (Jurden et al., 1996). In their study, Jurden et al. suggested that the different findings among the factor analytic studies outlined above could be the result of there being a close fit amongst all the competing models. Thus, whilst all models can provide a good fit for the data, Jurden et al. reported that the model that provided the best fit for the data included visual and verbal memory factors.

Results of factor analytic studies have consistently shown that the verbal memory tests load primarily on a general memory or a verbal memory factor (e.g. Larrabee & Curtiss, 1995; Leonberger et al., 1992). Measures of immediate visual memory have tended to show a strong association with visuo-spatial intelligence, whilst delayed recall had a stronger association with a general memory factor (for e.g., Larrabee & Curtis, 1995; Leonberger, Nicks, Goldfader & Munz, 1991; Leonberger et al., 1992).

The apparent variability in the results of these studies could be interpreted in a number of ways. For example, it is possible that factor structures are different for different populations. Indeed, Wechsler (1987) found different factor loadings across normal and clinical samples and Bornstein and Chelune (1989) found different patterns of factor loadings across different ages, educational background and intelligence levels. Psychometric considerations may also influence the results obtained, including the method of Factor Analysis, the criterion chosen to retain factors, sample characteristics, the reduced reliability of some of the subtests and the potential for measurement covariance to complicate results when immediate and delayed measures were included (Elwood, 1991; Loring, Lee, Martin & Meador, 1988; Smith et al., 1992a). Differences across studies in these areas make any comparison between studies difficult. In factor analytic research, the labeling of factors is also arbitrary (Heilbronner, 1992). Thus, based on factor loadings, different research groups may assign different interpretation and labels to factors.

Although there have been discrepancies in reported factor structures, confirmatory factor analytic research has suggested that the WMS-R is likely to contain at least three components including attention, immediate memory and delayed memory. However, this factor pattern may vary across clinical groups and nonverbal and verbal memory factors may be present in some groups, particularly in those with unilateral brain damage (Bornstein & Chelune, 1988; Bowden, 1997).

Wechsler Memory Scale- Third Edition.

The WMS-111 indexes were said to be constructed based on meaningful aspects of clinical memory assessment (Wechsler, 1997). Confirmatory factor analysis of the WMS-111 resulted in a five factor solution. There factors included attention/ concentration, immediate auditory memory, immediate visual memory, delayed auditory memory and delayed visual memory (Wechsler, 1997). However, closer inspection of the results indicated that the model supporting the index structure was not significant in the lower age group (30-64) and only approached significance in the 65-89 year age group. Thus, the factor structure could change across age. The intercorrelations with some of the subtests also would suggest that the five factor structure may not be stable across all age groups. For example the visual subtests do not correlate highly and factor analytic techniques are based on correlation. Further factor analytic studies in a range of clinical and age groups are required to investigate the underlying structure of the WMS-111.

Diagnostic Utility

For a test to be both valid and clinically useful, it needs to be sensitive to different types of memory disorders and be able to differentiate between different groups of brain damaged individuals. With regard to the WMS-R, studies have shown that the difference between the General Memory and Delayed Memory Indexes can distinguish between amnesic, demented (Alzheimer's and Huntington's) and control participants (Butters et al., 1988). Relative to normal controls, scores on all five indexes has been found to be poorer in recently detoxified alcoholics (Ryan & Lewis, 1988), head injured individuals (Reid & Kelly, 1993), and individuals with multiple sclerosis (Fischer, 1988). Troster et al. (1993) also found that savings scores derived from the Visual Reproduction and Logical Memory subtests differentiated between normal controls and individuals in the early stages of Alzheimer's Disease and Huntington's disease. These results suggested that the scale was sensitive to brain injury and could discriminate between some clinical groups.

Studies have also investigated whether the WMS-R is sensitive to the effects of unilateral brain lesions. Comparison of these types of individuals on the Verbal and Visual Indexes has provided mixed results. Generally, studies have not found that the Visual Index is more impaired in samples with right hemisphere damage. However, left hemisphere damage has been found to result in a reduction on the Verbal Memory Index (Wechsler, 1987; Chelune & Bornstein, 1989; Loring et al., 1988).

Performance on the Verbal and Visual Indexes has been found to incorrectly predict unilateral temporal lobe damage (Loring et al, 1989b). Chelune and Bornstein (1988) also found that groups with left and a right hemisphere damage were not significantly different on the Visual and Verbal Indices. However, they did find that those with left hemisphere lesions were more adept at learning and retaining nonverbal/visual material than comparable verbal material. They also found that the right hemisphere group was more adept at learning verbal than non-verbal material. Bornstein et al. (1988) found that a group with right hemisphere lesions only displayed a non-verbal memory deficit when proportion recall scores derived from the Visual Reproduction and Logical Memory subtests were compared.

These findings have not provided conclusive support for the Visual and Verbal Indexes being able to discriminate between unilateral lesion groups. However, the Visual and Verbal Indexes were based solely on the immediate recall portions of the test and therefore they may not have been sensitive to difficulties retaining information over time.

Preliminary studies of the diagnostic utility of the WMS-111 were presented in the technical manual and suggested that the test had potential to discriminate between normal controls and a number of clinical groups. Small samples of individuals with Alzheimer's Disease, Huntington's Disease, Parkinson's Disease, closed head injury, Multiple Sclerosis, Temporal Lobe Epilepsy, Korsokoff's Syndrome, Chronic Alcohol Abuse were administered the WMS-111. Generally, the results of these groups were consistent with the predicted expectation. For example, those with Alzheimer's Disease were significantly impaired on all Indexes, apart from Working Memory. Those with Huntington's Disease had significantly higher retention and retrieval composites compared with the Alzheimer's Disease sample. The Huntington's Disease group was also disproportionately aided by recognition. Given the small sample sizes, further independent studies are needed to investigate the diagnostic utility of the WMS-111 indexes and subtests.

The following section will investigate the validity and reliability of several of the subtests that have been present on the WMS and its revisions.

Content Validity of Selected Subtests

Logical Memory

Wechsler Memory Scale.

The Logical Memory subtest from the original WMS did not include a delayed recall procedure in its original form and the scoring guidelines for the subtest were poor. This limited the clinically meaningful information that could be obtained from it. The subtest was improved on the WMS-R by including a delayed recall trial and by adding a more detailed scoring system.

Wechsler Memory Scale- Revised.

The Logical Memory subtest on the WMS-R had good reliability and good criterion validity. In factor analytic studies, the subtest consistently loaded with other memory tasks on a separate factor from tests measuring verbal comprehension (Larrabee & Curtis, 1995; Leonberger et al., 1992; Nicks et al., 1992). Deficits on the Logical Memory task have also been found in samples with left hemisphere compromise (Bornstein et al., 1988; Naugle, Chelune, Cheek, Luders & Awad, 1993).

Despite its general robustness as a measure of verbal memory, the subtest format on the WMS-R was criticised by Loring (1989a) for not providing any cued/recognition format, although prompts were allowed in the delayed recall condition. However, the type of prompt used could bias retrieval (Loring & Papanicolaou, 1987). To control for this potential bias in retrieval a multiple choice format may be more appropriate. Fastenau (1996) developed a multiple-choice procedure with 20 questions asked about each story. These questions identified the main story units. Results suggested a normal distribution in a less educated group, but a mild ceiling effect in a more educated group. The procedure was potentially very time consuming and participants may have been asked questions about story items that they had already freely recalled. Thus, although this was a potentially useful addition, some adjustments could be made to make it more efficient and clinically useful.

Wechsler Memory Scale- Third Edition.

The WMS-111 retained the Logical Memory subtest with some content and administration changes. Story A was retained, but Story B was replaced with the rationale that the new story had less emotive content. However, Story A could also be considered as having emotive content. The gains of replacing Story B may not outweigh the costs of the difficulty in comparing performance across the WMS-R and WMS-111 versions of the subtest. The administration of the subtest was also changed, with Story B being presented twice and recall being tested immediately after each presentation. The repetition of Story B could enable those with poor memory to learn more material and in turn possibly facilitate some retention of information over time. However, it is a new procedure that was not based on any established methodology of memory assessment or model of memory functioning. Thus, interpretation of performance on this subtest could be problematic.

By presenting Story B twice, retroactive interference effects from Story B on Story A may be enhanced. When scoring, the three immediate recall scores are summed and no normative data was presented in the manual on the pattern of recall in each condition. Thus, an evaluation of the effects of interference cannot be made against a normative standard. No adjustment in scoring is made when combining delayed recall scores for stories A and B for the fact that Story B was presented on two occasions. Thus, this delayed recall score could be weighted with recall of Story B. These changes could limit the interpretations that can be made about recall and any comparison between the performance on the WMS-R and WMS-111 would be difficult.

A potentially positive addition to this subtest included scoring of thematic content as well as the traditional scoring approach. This addressed the criticism that requiring verbatim responses did not reflect the "normal" processing of memories, that is, extracting essential features and then transforming and incorporating this information into an existing knowledge (Loring & Papanicolaou, 1987). Thus, a scoring system based on main themes may more adequately reflect everyday memory processing rather than one based on word for word recall. The reliability of the procedure was found to be sound (Immediate .77; Delayed .79) although test-retest reliability was poorer (Immediate .64; Delayed .68 for ages 55-89).

A further addition to the Logical Memory subtest was the assessment of delayed recognition after delayed recall. This involved administering 30 yes/no questions that focused on the details of the stories. This format could be potentially time consuming and frustrating for individuals. As it was a forced choice task, it could also be difficult to distinguish true recognition from guessing and there were no repeat questions to evaluate recognition consistency. Again, despite Story B being presented twice, there was no distinction made between the recognition of the two stories. Thus, the recognition procedure may be of limited use because of the difficulties inherent with interpretation of scores.

Psychometrically, the Logical Memory subtest had highly stable reliability coefficients across the age groups of the normative sample (Average Immediate .88; Average Delayed .79). Test-retest reliability was strong (Average Immediate .80; Average Delayed .76 for ages 55-89) and interscorer reliability was reported to have been greater than .90. It had high a correlation with the Verbal Paired Associate subtest (Average Immediate .48; Average Delayed .46), both subtests contributing to the Verbal Indexes. This suggested that these two tasks shared some common underlying functions. The subtest had uniformly low correlation with the Faces and Visual Reproduction subtests, both visually presented.

The Logical Memory subtest had a moderately high correlation with the Family Pictures subtest, the latter being visually presented (Average Immediate .40; Average Delayed .46). This would suggest that the Family Pictures subtest had at least a moderate verbal loading, which is consistent with the material being highly verbalisable and the response mode being verbal.

Although the Logical Memory subtest would appear to have retained its psychometric robustness, the changes to the subtest in the WMS-111 have considerable implications for clinical practice. Administration time is likely to be longer, any retroactive interference effects cannot be adequately inferred from the scoring method and the recognition measure is flawed. Thus, the changes to this subtest in the WMS-111 may detract from the clinical information that can be obtained.

Verbal Paired Associate Learning

It has been postulated that verbal paired associate learning tasks represented one of the most sensitive instruments to assess memory impairment (Erikson & Scott, 1977; Kapur, 1988). An association between the left temporal lobe and paired associate learning, particularly with unrelated word pairs, has been found in participants temporal lobotomies and neuronal loss/sclerosis (Frisk & Milner, 1990; Rausch & Babb, 1993; Saling et al., 1993). These results suggest that Verbal Paired Associates tasks are valid measures of verbal memory impairment.

Wechsler Memory Scale (WMS).

The Associate Learning subtest that appeared on the WMS did not include a delayed recall trial, limiting the information obtained about memory and there was some criticism about the inclusion of "familiar" pairs. The clinical usefulness of the standard version was therefore likely to be reduced.

Wechsler Memory Scale- Revised (WMS-R).

The Verbal Paired Associates subtest on the WMS-R was slightly modified from the original but still included novel and familiar word pairs. It also included a delayed recall trial.

The inclusion of familiar word pairs in associate learning tasks has been an area of debate, for example, Leng and Parkin (1990) suggested that using familiar pairs was uninformative as they could be guessed. There has also tended to be a ceiling effect in a younger normal population when using familiar pairs. The use of only unfamiliar or "novel" pairs has been supported by research that has indicated that the familiar pairs and unfamiliar pairs load on separate factors in factor analytic research (Oresicke & Broder, 1988). Studies have also demonstrated that unfamiliar pairs correlate more highly with other measures of verbal memory (MacCartney-Filgate & Vriezen, 1988), that they are more clinically sensitive than familiar pairs (Fischer, 1988; Trahan, Larrabee, Quintana, Goethe & Willingham, 1989) and that they are more closely associated with left hippocampal dysfunction (Rausch & Babb, 1993; Saling et al., 1993).

Wechsler Memory Scale- Third Edition (WMS-111).

Perhaps in response to research findings, the Verbal Paired Associates task on the WMS-111 only included "novel" word pairs said to be of "high imagery". This subtest had very good reliability (Average Immediate .93; Average Delayed .83) and strong test-retest reliability (Average Immediate .83; Average Delayed .79 for ages 55-89). It also shares a high correlation with the Logical Memory subtest as would be expected and a low correlation with the Faces subtest and Family Pictures (Average Immediate.34; Average Delayed .36).

By removing the "familiar" pairs from the paired associate task, the clinical diagnostic benefit of the familiar/novel pairs is likely to have been lost when using this version. In particular, the difference between new learning and "refreshing" of old associations will be lost. The use of "familiar" word pairs can also be very informative in a population with memory impairment. Although paired associate tasks based on only "novel" pairs may eliminate the ceiling effects that have been found in some younger and better-educated groups on the WMS-R, these tasks may be to difficult for the elderly and some clinical groups resulting in floor effects. The use of only "novel" pairs may also be better able to better distinguish those with left hippocampal versus left neocortical damage (Saling et al., 1995). However, this is often not the only goal of clinical memory assessment. Thus, there are advantages and disadvantages to including or excluding familiar word pairs.

To overcome the dilemma of including or excluding familiar word pairs, paired associate tasks could include them, but have some scoring modification. If the scoring system involved separately collecting scores on novel and familiar pairs, then a comparison between the two (familiar and novel) could provide clinically meaningful information. In addition, the recall of novel pairs could still be used for distinguishing verbal memory impairments in those with left hippocampal damage.

Despite the concern over the omission of "unfamiliar" pairs, the inclusion of a delayed recognition task on the WMS-111 was a potentially positive addition. However, like the other attempts in the WMS-111 to include recognition testing, there were a number of problems with the procedure.

Firstly, as recognition was tested via a yes- no response, guessing could confound true recognition. If some control for response bias was incorporated into the scoring procedure, this problem could be reduced. In fact, four of the pairs were tested twice during the recognition test, which could have permitted some check of consistency. However, the manual only gave norms for the total score and the total score was in fact biased *because* the recognition task tested four pairs twice. This meant that the total score could be substantially influenced by which pairs were committed to memory.

Secondly, administration could be time consuming, particularly for an older population, given that 24 word pairs are presented. Thirdly, recognising either the first word or the second word of the pair was sufficient for a correct response. Thus, memory for a word, rather than memory for the association was being assessed in this task. Finally, recognition was tested straight after delayed recall, when the participants had just had the first word of each pair presented to them. Thus, recent memory for either the first or the second word of a pair could confound this "recognition" measure. Overall, the recognition procedure was significantly flawed and so the data derived was questionable.

Faces

Memory tasks using faces appear to have strong ecological validity for non-verbal memory assessment (Loring & Papanicolaou, 1987; Mayes, 1986). A number of studies, including visual matching of unfamiliar faces, processing of faces and identification of famous faces have suggested that the right hemisphere is specialised for processing faces (Carlesimo & Caltagirone, 1995; Newcombe, de Haan, Ross & Young, 1989). Studies have also shown that memory for faces is sensitive to right hemisphere deficits (Morris et al., 1995b; Naugle, Chelune, Schuster, Luders & Comair, 1994). Given the material specific model of memory and the association between the right hemisphere and non-verbal memory, these findings suggest that tasks using faces measure some aspect of non-verbal memory. Because faces are rich in detail, they are likely to reduce the possibility of rapid verbal encoding and are more likely to require the persistence of a detailed visual image (Milner, 1968). Thus, tasks using faces are likely to be good tests of non-verbal memory and be sensitive to the effects of right temporal lobe resections. Despite this, there are few clinical tests assessing memory for faces.

Warrington's (1984) Recognition Memory Test (RMT) incorporated a task assessing memory for faces via a recognition format. Studies have demonstrated that the RMT is sensitive to lateralised temporal lobe neoplasms and infarctions (Warrington, 1984) and right temporal lobectomy (Herman et al, 1995a; Morris et al., 1995b). However, it was not found to be sensitive to right temporal lobe epilepsy (Baxendale, 1997; Herman et al, 1995a).

Although the RMT appeared promising as a measure of verbal and non-verbal memory, there were some criticisms, including: 1) low ceilings 2) a high percentage of false positives, 3) no reliability estimates, and 4) the possibility that the faces could be easily verbalised due to the photographs including features such as clothing and hairstyles (Adams, 1989; Kapur, 1987). In addition, because a test of free recall was not included, a direct comparison between the integrity of encoding and retrieval could not be made.

A face memory task has also been included on the WMS-111. Despite its apparent "ecological validity", there are also some potential problems with the task. Firstly, slow rate of information processing or sensitivity to interference may depress scores. As the faces were only presented for a two seconds, those with additional cognitive problems may do poorly on this task, unrelated to their actual memory capacity. Scoring of the task was also simplistic, with only the number of correct yes responses being collated. There was no adjustment made in the scoring for the decision criterion (percentage of yes to no responses) used by the participant or for variation in this criterion between immediate and delayed testing. Furthermore, there was no adjustment for response bias, that is, responding with only yes or only no responses. Thus, although the use of faces as stimuli in a memory task was a potential positive admission, the procedure and scoring were likely to limit this potential.

The Faces subtest also had one of the lowest reliabilities of all subtests on the WMS-111, ranging from .65 to .80 for the immediate trial and .66 to .83 for the delayed trial. It also had one of the lowest test retest reliabilities on the WMS-111. This suggests that it may not be a useful tool to interpret independent of other results and that confidence intervals for scores should be taken into account when interpreting performance.

The Faces subtest would also appear to be relatively independent of the other subtests on the scale. The subtest had a low correlation with the Logical Memory subtest (Average Immediate .14; Average Delayed .22) and the Verbal Paired Associates subtest (Average Immediate .18; Average Delayed .22). This would suggest that it may be testing a different aspect of memory functioning (or at least a different cognitive function) than verbal memory. A possible interpretation would be that it draws on non-verbal memory given past research that has shown an association between the right hemisphere of the brain and performance on face memory tasks.

The Faces subtest did not have a strong correlation with the other visually presented subtests in the WMS-111. The Faces subtest had only a modest correlation with the Family Pictures subtest (Average Immediate .30; Average Delayed .28), although the correlation tended to be stronger in the elderly. Interpretation of the Visual Immediate and Visual Delayed Indices could be problematic because these two subtests do not share a strong correlation. Although these were both visually presented subtests, it would appear that Family Pictures draws on a very different aspect of memory than the Faces subtest.

The correlations between the Faces and the Visual Reproduction subtests, both immediate (range=.18-.23) and delayed (range=.15-.27) were also quite low. Thus, although they were both presented visually it would appear that they were drawing on different processes. However, potential difficulties with the Visual Reproduction subtest could detract from its ability to measure non-verbal memory. In particular, it may be that the scoring of the Visual Reproduction subtest of the WMS-R is not sensitive to non-verbal memory functioning.

Visual Reproduction

Wechsler Memory Scale (WMS).

The WMS version of Visual Reproduction subtest did not include a delayed recall trial and scoring rules were not explicit, making it difficult to evaluate the reproductions. In factor analytic studies, the subtest loaded primarily on a visual-perceptual motor ability factor (Heilbronner, Buck & Adams, 1989; Larrabee, Kane & Schuck, 1983; Larrabee, Kane, Schuck & Francis, 1985) raising questions about the validity of this test as a measure of memory and/or visual memory.

Wechsler Memory Scale- Revised (WMS-R).

The Visual Reproduction subtest of the WMS-R incorporated some positive changes from the original scale. Explicit scoring criteria were included, resulting in better scoring reliability (Loring & Papanicolaou, 1987). The addition of a delayed recall procedure enabled a measure of retention of information over time to be calculated. However, several criticisms were directed at this subtest.

Firstly, recalling visual designs requires adequate visual perceptual and constructional skills and sufficient motor functioning. As a result, scoring and interpretation of poor performance could be confounded by constructional dysfunction, perceptual deficits or poor motor abilities (Gfeller, Meldrum & Jacobi, 1995). Indeed, Gfeller et al. found that individuals with constructional dysfunction exhibited impaired performance on immediate and delayed recall of the designs as compared with normal participants. Likewise Haut, Weber, Wilhelm, Keeover and Rankin (1994) found that problems in perception and construction skills affected performance on the Visual Reproduction subtest in a group with Alzheimer's Disease. Ricker, Keenman and Jacobsen (1994) also found that participants with vascular and Alzheimer's dementia who had poor visual perceptual skills also had a visual memory deficit as reflected by their significantly reduced performance on memory for designs.

However, their poor visual skills did not fully explain their visual memory deficit. Further evidence of the large constructive, perceptual and motor components involved in this task has been illustrated in factor analytic research. Many of these have demonstrated a strong relationship between immediate recall of visual reproductions and visual-perceptual-motor tasks (Bornstein & Chelune, 1988; Larrabee & Curtis, 1995; Leonberger et al., 1991, 1992).

The strong correlation between the Visual Reproduction subtest and tasks involving visual constructional, perceptual and motor responses does not preclude it having a valid role in the measurement of non-verbal memory (as compared to cognitive skills). Memory is a higher order cognitive process and therefore performances on tests of memory are likely to reflect the integrity of lower order cognitive processes. Nevertheless, it is important for the clinician to ascertain whether poor performance on a non-verbal memory task is due to memory impairment rather than another neuropsychological deficit. Thus, when administering a memory task, either non-verbal or verbal, it is important to examine component processes to rule out a deficit in other areas affecting performance on the memory task.

In order to examine the component processes contributing to performance on non-verbal memory tasks, procedures need to be incorporated into test design to enable deficits in visuo-perceptual, visuo-constructional and visuo-motor abilities to be separated from memory performance and possible non-verbal memory deficits. For example, a recognition trial could be used to assess visual memory without visuo-constructive functions (Boller & de Renzi, 1967), a copy trial could be used to identify participants with visuo-constructive and visuo-spatial problems (Haut, Weber, Demarest, Keeover & Rankin, 1996) and a matching trial could be incorporated correct for the effects of poor perceptual skills (Haut et al., 1994).

Fastenau (1996) devised recognition (multiple choice), matching and copy trials for the Visual Reproduction subtest. He found that the new measures had limited reliability, poor psychometric properties and ceiling effects (particularly in high education and low age groups) as an artifact of the small number of items on recognition and matching. Despite the psychometric difficulties, Hanger, Montague and Smith (1991) found that the recognition trial did differentiate between neurological and non-neurological groups. These preliminary studies suggest that the addition of a recognition procedure is potentially of considerable value.

Poor performance on a copy trial may suggest difficulties with visual construction. However, by itself, a copy trial does not provide a quantitative evaluation of the contribution of construction and memory to performance. Haut et al. (1996) developed a method for computing the contribution of visual construction to recall of designs. After the standard administration of the subtest, they asked each participant to copy the designs. They then divided the subtest raw score for both the immediate and delayed conditions by the total score obtained by copying the design. This proportion score was then used to quantify memory performance independent of visuo-constructional skills. When using the proportion score, they found that a group with Alzheimer's disease had a greater non-verbal memory impairment on the subtest as compared to control participants, independent of their constructional skills. Thus, additions to the subtest, such as a copy trial, may allow more clinically useful information to be obtained.

The Visual Reproduction subtest has also been criticised because the simplicity of the stimuli could result in some participants being able to verbally encode the designs. In fact, the scoring system for this subtest in the WMS-R version was based on verbal criteria. It is therefore possible that the designs can be processed by the left rather than right hemisphere (Chelune & Bornstein, 1988; Heilbronner, 1992). If this is the case, the test may no longer measure the construct of non-verbal memory, but perhaps a combination of skills and abilities.

To overcome the problem of verbal encoding, Lee, Loring and Thompson (1989) have suggested that a non-verbal memory test needs to use complex and unfamiliar stimuli that are very difficult to encode verbally. However, complex, random shapes may be too complex for *both* normal *and* brain damaged individuals to recall and even complex visual material may be verbalisable, at least to some extent (Cermack & Tarlow, 1978; Eadie & Shum, 1995; Vanderplass & Garvin, 1959). Thus, it may be difficult to develop stimuli that cannot be verbally encoded by some people and that are not too difficult for individuals to recall.

It is possible that individuals normally use verbal encoding strategies when processing visual information. Therefore, a test that eliminates the use of this strategy may not provide relevant information about real-life memory impairments (Heilbronner, 1992). Thus, devising "verbal" free non-verbal tests may be a goal of research into neural substrates of memory, but not clinical memory testing.

An alternative to the use of complex figures is to acknowledge that current non-verbal memory tasks have a verbal component and are likely to be processed verbally and non-verbally. It may be counter-intuitive to expect individuals to switch off verbal processing when confronted with non-verbal material. Because verbal encoding may assist on the Visual Reproduction subtest, those with left hemisphere damage may also do poorly because of deficits in verbal encoding and memory. Indeed, Trahan, Quintana, Willingham and Goethe (1988) found that a group with left hemisphere strokes exhibited deficits on the Visual Reproduction subtest. Some non-verbal memory and verbal memory tests also correlate moderately, suggesting that they may both be partially mediated by similar processes (Larrabee & Curtiss, 1995).

Despite this, clinicians often wish to understand the relative contribution of defective verbal or non-verbal memory. As currently administered and scored, there is no way for the clinician to quantitatively extract the role of verbal and non-verbal encoding processes to a person's recall on the Visual Reproduction Subtest. One option to address this difficulty includes devising a new test that includes complex stimuli. Alternatively, the scoring criteria of an existing test, such as Visual Reproduction, could be redeveloped to focus on those aspects of the designs that are more likely to be non-verbally encoded. Thus, instead of trying to eliminate what may be a natural mode of memorising for some people, the scoring system of the task could be altered to focus scoring on the nonverbal aspects of the designs. In this way, information could be gained about the functional deficit to the individual (the total score) and the extent of their nonverbal memory difficulties (the score on the non-verbal items) without changing the content of the task.

Several criticisms can be directed at the standard Visual Reproduction scoring criteria. The standard scoring system included differential weighting of each of the four designs. There was no rationale or logical basis presented for this differential weighting in the WMS-R manual. Although it is possible that the designs contain an unequal amount of elements and therefore differ in complexity, no data concerning this possibility was provided in the WMS-R manual. The problem with weighting the designs is that if a participant performs badly on one design (for example Design D), for some reason other than memory such as attention, their score could be unduly depressed. If the items were equally weighted, and performance on one item was contaminated, then there would be a more logical basis for estimating the overall performance by using the scores on the other three designs. The scoring system may be better structured if each design was given the same weighting so that if performance on one design was contaminated a prorata score could be obtained.

Allocating an equal number of points for each of the designs would also create a potential to build a database of differential performance on each design. Studies using the standard scoring system have typically reported on the total score rather than the score on individual items. Given that the items are not equally weighted it is difficult to make a qualitative evaluation of what, if anything, is the significance of performing differently across the designs. By weighting each design equally there would be the potential to compare the performance across the designs in different clinical groups and evaluating what the importance of any difference may be. In the long term, studies could investigate whether there were patterns of performance occur across individual designs and whether this provided diagnostic and clinical information.

Not only did the scoring system include an unequal number of points for each design, the range of scores was restricted. This was particularly true for the first three designs on the WMS-R (Designs A and B contain 7 items and design C contains 9 items) that collectively contributed only 54% to the total score. This small range potentially reduces the discriminatory power of the subtest, particularly at lower and higher levels of performance. Restricting the range of scores could also compound the low reliability of the subtest.

In the standard scoring system, items included little allowance for carelessness in drawing, poor drawing ability or poor visuo-motor coordination. For example, in Design Three, if an individual has drawn 15 dots and one circle (perhaps carelessness) they receive zero for this item. However, the individual would have clearly retained the idea that each quadrant had four round elements. This lack of flexibility in scoring could place pressure upon the clinician to make a judgement as to whether they should score the product (15 dots and 1 circle) or the intent (16 dots) of the reproduction. Rather than penalising occasional lapses (for e.g. forgetting one dot) or poor motor control or drawing ability (for e.g. having gaps between the join of lines and overshoots across lines) a scoring system should have greater tolerances so that memory recall, rather than precision in drawing is rewarded. This change could also improve the reliability in scoring individual items, because clinicians would not be placed into the position of questioning if they should score the product or the intent.

The standard scoring system included many items that focused on measurement of precise angles and distances. It could be argued that because of this focus, the scoring system would penalise carelessness in drawing rather than "forgetting". For example, a low score on Design One could be due to poor motor control resulting in one line being longer than another line (Item 1), staffs not intersecting at midpoints due to the longer lines (Item 2) and flags not sharing a side with the staff (Item 6). However, this material may have been remembered correctly, but not produced accurately. If it is the reproduction of the essence of a design that is required, then some flexibility concerning drawing needs to be permitted in a scoring system.

Measurement of precise angles and lengths could potentially result in poor inter-scorer agreement. As measurement on the standard scoring system can be considered quite strict, the burden would be placed on the scorer to judge whether to score for accuracy or intent. Thus, the scorer would need to decide whether to penalise for poor reproduction of angles, even when it was clear that there had been some memory recall. It could therefore be argued that a scoring system should include greater tolerances for the measurement of angle degrees and the length of lines. By increasing the tolerances allowable, scoring would be less dependent on precise drawing and focus on memory recall. Scoring would also potentially be more objective, because scorers would be less likely to make a decision on scoring detail versus intent.

There are many instances in the standard scoring system where failure on one scoring criterion can preclude any score on other criteria. Thus, partial recall is not rewarded and an individual's recall could be under-estimated by the scoring system. In clinical practice, participants are often noted to recall partial elements of a design, but not receive any credit for this. For example, in Design Three, participants often only include a horizontal *or* vertical division (Item 2). It could be argued that a scoring system should include a number of items that address both partial and total recall. The inclusion of more lenient items addresses the problem of a participant not receiving credit for incomplete or partial recall. However, in order to discriminate the quality of recall, stricter items could be included that require exact reproduction.

When applying the WMS-R scoring system, not all items are applied independently. As a result, a failure to meet the criteria for one item results in a zero score for other items. For example when scoring Design 3, if the participant produces fewer than four squares (item 5), items 6-9 are also scored zero. However, if three squares are produced, this clearly indicates some degree of memory recall. If the three squares are square in shape (Item 6), bisected by vertical and horizontal lines (Item 7), not rotated (Item 8) and in equal size and proportion (Item 9), then considerable recall has been penalised because there were three, not four squares. Although three squares may represent a degree of memory lapse, clearly not receiving a score for any of these items underestimates an individuals' recall. Thus, a score may not adequately reflect the level of memory recall because scoring some items is dependent upon receiving a score on another item.

Linking items in this way, for example making a score for Item 8 on Design Three dependent on a score for Item 5 could also be understood as implying that information is retained in a certain way. For example, it could be seen as implying that memory for the bisecting lines (Item 7) is dependent upon memory for the four squares (Item 5). Clinical data does not support this, with many clinicians having assessed the participant who draws bisecting lines in the four quadrants, without any internal squares. Thus, there may be little basis for making items dependent upon one another, and certainly there was no rationale outlined in the WMS-R manual. Mindful of these concerns, a scoring system could be developed that involved scoring all items independently. In this way, the final score would make no assumption about the way information was retained and the lack of recall for a particular item would not, on its own, have a significant impact on the score. In addition, concerns about a score underestimating memory recall would be reduced, because a score on one item would be dependent on recall of another item.

The standard Visual Reproduction scoring system was based on semantic cues, which emphasised that the material could be verbally encoded. Indeed, a major criticism of the Visual Reproduction subtest has been that the stimuli were potentially verbally encoded (Eadie & Shum, 1995; Chelune & Bornstein, 1988; Heilbronner, 1992; Lee et al., 1989). Rather than focussing on attention to detail, the scoring system could focus less on precise scoring of angles and focus more on those aspects that are more likely to be non-verbal encoded, such as the relationship between easily verbalised shapes. It may also be that a certain subgroup of items on the scoring system may be a better indicator of non-verbal memory. The use of an Index of items to represent non-verbal memory performance, rather than the total score, would acknowledge that some verbal encoding is likely to take place on this task.

A number of the limitations/problems with the standard scoring system could lead to reduced scoring reliability. The inter-rater reliability of the scoring system has been examined by a number of authors (McGuire & Batchelor, 1998; Wechsler, 1987; Woloszyn, Murphy, Wetzel & Fisher, 1993). In the original standardisation sample of the WMS-R, Wechsler (1987) reported inter-scorer reliability coefficients of .97 for the Visual Reproduction subtest. Similarly, Wolozyn, Murphy, Wetzel & Fisher (1993) reported reliability coefficients of .977 for immediate recall and .975 for delayed recall in a clinical population. McGuire and Batchelor (1998) reported moderately strong Pearson's coefficients of .82 for immediate recall and .94 delayed recall when examining inter-rater reliability in a clinical population. However, reliability coefficients on individual Designs ranged from .49 (Design One Immediate Recall) to .94 (Design Four Immediate Recall).

Although a number of studies have found high levels of inter-rater reliability when examining correlation coefficients, raw score differences between scorers on individual designs has been found to be large in some cases. Wechsler (1987) reported that on each of the four designs, raters differed by up to four points. This could equate to a difference between raters of 16 points across the four designs, although no information was provided on how frequently this occurred. However, Wechsler (1987) did report that the average absolute difference in raw scores for total scores was 1.5 for Visual Reproductions, suggesting that large differences only happened rarely. McGuire and Batchelor (1998) also reported that the score on each individual stimulus card could vary by up to four points between raters, although the overall mean difference between scorers was only four points. Five of their forty participants had total scores that differed by 10 points or more between the two scorers, again suggesting that large differences did occur but were relatively infrequent.

The Visual Reproduction subtest of the WMS-R was included in the battery as a measure of non-verbal memory. As outlined, the actual scoring of the reproductions may influence the subtests ability to measure non-verbal memory. Indeed, support for the subtest as a measure of non-verbal memory has been inconclusive. Factor analytic studies have consistently indicated a primary loading for immediate recall on a visual-perceptual- motor ability factor and only secondarily on memory (for e.g., Bornstein & Chelune, 1988; Larrabee & Curtis, 1995). This loading with spatial tasks in factor analysis studies suggests that the immediate recall portion is more highly associated with visuo-constructive and visuo-perceptual skills than with memory. Whilst, delayed recall has tended to load on a general memory factor, it still retains strong associations with other visuo-constructive based tests (for e.g., Larrabee & Curtis, 1995; Leonberger et al., 1991, 1992).

Evidence from lesion studies has also been inconclusive with respect to the association between memory for Visual Reproductions and the right hemisphere of the brain. Contrary to expectations, poor performance on Visual Reproduction has not been associated with right hippocampal volume (Bigler et al., 1996; Sass et al., 1992) and has been related to left hippocampal volume in some groups (Bigler et al.).

Perhaps of more importance clinically, the test has not been consistently found to be sensitive in detecting poor memory performance in persons with right temporal lobe lesions (Chelune & Bornstein, 1988; Lee et al., 1989; Loring et al., 1989b; Naugle et al., 1993; Sass et al., 1992). However, delayed recall on the Visual Reproduction subtest has been shown to be superior to immediate recall for discrimination of right unilateral temporal lobe seizure activity by some researchers when percent retention has been used as the measure of memory (Delaney, Rosen, Mattson & Novelly, 1980). In contrast, Bornstein et al (1988) found that a right temporal lobectomy group did not perform significantly worse on the their actual Visual Reproduction test score, although the difference approached significance. However, they did find that those with right hemisphere damage were more adept at learning verbal rather than non-verbal material. Chelune and Bornstein (1988) also found that the left hemisphere group did more poorly on verbal than non-verbal tests, and the right hemisphere group did more poorly on non-verbal as compared to verbal tests. However, their performances only differed significantly on the verbal subtests (Logical Memory and Verbal Paired Associates).

These findings, that poor performance on the Visual Reproduction subtest is not consistently associated with right hemisphere dysfunction, could be related to the participant characteristics in these studies. For example, individuals who have temporal lobe epilepsy are not a homogeneous group. These individuals can differ in the extent of their memory difficulties pre-operatively. Those who have undergone temporal lobectomy may also vary considerably in their post-operative outcomes. Some individuals demonstrate a clinically significant memory impairment whilst others show and report no memory impairment (Baxendale, 1997; Herman et al., 1992; Powell, Polkey & McMillan, 1985). Indeed, it has been found that the adequacy of post-operative memory can be related to many factors, including the extent of the hippocampal resection (Jones-Gottman, 1987; Smith & Milner, 1989), preoperative hippocampal volume (Trenerry et al., 1993), older age of onset of epilepsy (Herman, Seidenberg, Haltiner & Wyler, 1995b; McMillan et al., 1987; Powell et al., 1985; Saykin et al., 1989; Trenerry et al., 1993), older chronological age (Helmstaedter & Elger, 1996; Herman et al., 1995b), and the adequacy of preoperative memory (Chelune et al., 1991; Helmstaedter & Elger; 1996; Helmstaedter , Hufnagel & Elger, 1992; Herman et al., 1992, 1995b; Herman, Wyler & Somes, 1993; Powell et al; 1985; Trenerry et al., 1993).

Given the range of factors that can potentially influence both pre-operative and post-operative memory performance, combining participants into groups may mask non-verbal memory impairment in those who do demonstrate post-operative decline. Subgroups may need to be further refined in order to find material specific decline in memory performance on the Visual Reproduction subtest and indeed other memory tasks.

As has been illustrated, several studies with the Visual Reproduction subtest have raised questions about what it measures and whether this is some aspect of non-verbal memory. The delayed recall portion of the Visual Reproduction subtest on the WMS-R may be a measure of general memory and poor performance on this task may be related to impaired functioning in the right hemisphere of the brain in some cases. However, it cannot be conclusively stated that Visual Reproduction, as currently given and scored, is representative of non-verbal memory functioning in all participant groups.

Wechsler Memory Scale- Third Edition (WMS-111).

Visual Reproduction was designated an optional subtest in the WMS-111, although the manual did not outline a rationale for this. The new task consisted of 5 designs, Item 1 and Item 5 being added, and Item 2 from the WMS-R being omitted. The scoring system was changed to allow 2 and 1 point response for most of the scoring criteria, raising the maximum score 104. Delayed recall was measured 25-35 minutes later followed by delayed recognition. The recognition task involved the individual being shown a series of designs and asked to identify the ones previously shown. A copy and matching trial were also included with the latter involving the individual being asked to match the original stimulus to one of six design choices.

Reliability for this subtest was satisfactory for the immediate and delayed trials (Average Immediate .79; Average Delayed .77), but reduced for the copy (Average .73) and recognition (Average .75) trials. Test-retest reliability was also reduced (Immediate .71; Delayed .71 for ages 55-89). As previously identified, the subtest had a low correlation with the Faces subtest. The correlation between the Visual Reproduction subtest and Family Pictures ranged from .26 to .45, with the correlation being stronger in the older age groups. Thus, the visually presented material on the WMS-111 did not appear to share a high association with each other, at least not one that was greater than their association with verbal memory tests. With regard to the Family Pictures subtest in particular, the stronger association with the Logical Memory subtest than other visual measures, suggested that it was at best a measure of verbal and non-verbal memory abilities.

There were a number of potential problems with this revision to the Visual Reproduction subtest. For example, Design 1 is a line with a flag at each end and Design 2 is two lines, placed at right angles, with flags at each end. Thus, design one is actually half of design two, which may lead to confusion in the individuals' recall and hence recall may be difficult to score and/or interpret. There were also many problems with the delayed recognition procedure including:

1) The number of items made it potentially lengthy and time consuming;
2) The two designs in both items 4 and 5 were presented separately, which was potentially confusing, with participants not identifying the correct designs because they were not accompanied by their corresponding design;
3) The distracters were very similar to target designs, which meant partial recognition could to lead to an incorrect YES decision.
4) Because the distracter designs were similar, individuals could start to "Recognise" previously presented distracters.
5) Recognition was tested via a YES-NO response, which meant that guessing could confound recognition. Although there was an opportunity to check the score for consistency because each item was presented twice it was not utilised.

Although a copy trial was also included in the revision, no scoring procedure was presented to quantify the contribution of poor visuo-constructional performance to non-verbal memory. Interestingly, those items that were present across the editions (crossed staffs and large square) have contributed different weighting to the scoring system of the editions. For example, the crossed staffs and large square were worth 57% of the score for the WMS, 39% of the score for the WMS-R and 26% of the score for the WMS-111. There did not appear to be a rationale given for this change. The result of this difference was that scores across subsequent editions of the test would be difficult to compare.

Generally, the updated version of the subtest has many potential limitations, which limit interpretation of recall on the WMS-111 Visual Reproduction subtest.

Is there a Distinction between Verbal and Non-Verbal Memory?

A distinction between declarative memory for material that is non-verbal and for material that is verbal in nature has been postulated in the research literature (Milner, 1971). It has also been suggested that there are different neural substrates for the memory of verbal and visually presented material (Squire & Butters, 1992). In this view, verbal and non-verbal memory factors would be dissociable.

In order to examine the possible existence of separate verbal and non-verbal memory processes, studies have examined the association between the right and left temporal regions and performance on verbal and non-verbal memory tests. Whilst compromise in the structures in the left mesial temporal lobe has been shown to consistently impair verbal memory and learning (Giovagnoli & Avanzini, 1999; Herman, Connell, Barr & Wyler, 1995a) the converse finding, an impairment in non-verbal memory associated with right temporal damage, has not been found consistently with a number of studies failing to support the association (Barr et al., 1997; Dobbins, Kroll, Tulving, Knight & Gazzniga, 1998; Helmstaedter, Pohl &

Elger, 1995). However, the finding has been more consistent when delayed recall has been examined or when the task has involved stimuli that are difficult to verbally encode, such as abstract designs or faces (Giovagnoli & Avanzini, 1999; Gliebner, Helmstaedter & Elger, 1998; Glosser, Deutsch, Cole, Corwin & Saykin, 1998; Eadie & Shum, 1995; Morris et al., 1995a; Trenerry, Jack, Cascino, Sharbrough & Ivnik, 1996).

Inconsistencies in the findings across these studies could be due to a number of factors other than the lack of a distinction in the neural basis of verbal and non-verbal memory. As is outlined on pages 57 to 58, subject characteristics could significantly affect results in these studies. In addition, the characteristics of the verbal and non-verbal measures that were used could influence results. For example, a number of visual memory tasks require a visuo-motor response and depend on adequate visuo-perceptual skills. Thus, the scoring and interpretation of a number of non-verbal memory tests could be confounded by impairments in construction, reduced motor skill or visual perceptual disturbance (Heilbronner, 1992). If this is the case, then the resulting performance is not indicative of non-verbal memory.

In order to examine the existence of dissociable verbal and non-verbal memory factors, a number of factor analytic studies have also been conducted examining tasks of memory. The findings of these studies have also been inconsistent and there has been a considerable discrepancy among the reported factor structures. For example, pages 37 to 39 outlines the factor analysis studies conducted with the Wechsler Memory Scale and its revision. A variety of different results have been reported for factor analyses of the WMS-R, ranging from one factor (Elwood, 1991b, 1993) to three factor solutions (Roth et al., 1990; Woodard, 1993). Some of these studies have identified separate visual and verbal memory factors (Bornstein & Chelune, 1988, 1989; Jurden, Franzen, Callahan & Ledbetter, 1996; Larrabee & Curtiss, 1995).

Separate verbal and non-verbal memory factors have emerged in studies using younger participants (Bornstein & Chelune, 1989; Chelune & Bornstein, 1988), participants with well-defined unilateral lesions (Bowden, 1997), in a clinical group from an addiction recovery unit (Jurden et al., 1996) and in normal samples (Bowden et al., 1999; Jurden et al., 1996). Both the Jurden et al. and Bowden et al. studies found that solutions incorporating immediate memory and delayed memory factors fitted their data. However, both studies also reported that divisions with separate verbal memory and non-verbal memory factors provided the closest fit of the data.

Hunkin et al. (2000) also identified separate visual and verbal memory factors in their study of the Warrington Recognition Memory Test, the Wechsler Memory Scale- Revised and the Doors and People Test. However, their analysis identified a single recall factor and separate visual recognition and verbal recognition factors. Thus, they found only partial support for the verbal/non-verbal distinction. They concluded that separate visual and verbal recall factors may not have emerged because the visual and verbal material used in free recall tasks is often subject to dual encoding, that is, the material is both verbally and visually encoded.

As has been outlined previously (pages 38-39), the apparent variability in the results of these studies could be interpreted in a number of ways. Additionally, as Jurden et al. (1996) suggest, a number of competing models may provide an adequate fit of the data, but other solutions, for example, those incorporating a visual and verbal factor may provide the optimal fit. Furthermore, the previous concerns regarding the measurement of non-verbal memory with current tests and the potential for performance to be confounded with other factors (poor visuo-constructional/ perceptual skills) apply to factor analytic research.

Factor analysis of the WMS-111 standardisation sample has provided support for the verbal/non-verbal distinction (Wechsler, 1997). A recent factor analysis of the WMS-111 standardisation sample (using the primary subtests of the scale) identified a three factor solution, namely, working memory, auditory memory and visual memory (Millis, Malina, Bowers & Richer, 1999). However, the visual memory factor was identified by the authors as flawed, because the Faces subtest did not share sufficient communality with the Family Pictures subtest in the visual memory factor.

The extent to which verbal and visual memory dissociates and under what conditions that it occurs is unclear. However, conflicting research evidence does not necessarily lead to the conclusion that there is not a division between verbal and non-verbal memory. As has been outlined, a number of alternative explanations can account for the lack of double dissociation in temporal lobe epilepsy studies (see page 57-58) and the absence of separate verbal and non-verbal memory factors in factor analytic research (see pages 38-39). Thus, further investigation is required in order to establish the relationship between verbal and non-verbal memory. Research and clinical evidence does not conclusively point to a single memory system, therefore further research is required to understand the components of memory.

Non-Verbal Memory Assessment

In clinical practice, assessment has commonly focussed on the differences between verbal and nonverbal memory (Lezak, 1995). Research has also focussed on investigating the existence of a theoretical and clinical dissociation between verbal and non-verbal memory performance (Chelune et al., 1991; Eadie & Shum, 1995; Heilbronner, 1992; Helmstaedter et al., 1995; Saling et al., 1993). Thus, a reliable and well validated measure of non-verbal memory is an important aim in test development for both research and clinical reasons. The Visual Reproduction subtest has been commonly used as a measure of non-verbal memory (Lezak, 1995; McGuire & Batchelor, 1998). However, as has been outlined this task has a number of limitations as it is currently administered and scored.

It would appear that the standard features of a "non-verbal memory test" should include:

1) immediate, delayed and recognition trials to clarify the contributions of encoding, consolidation and retrieval;
2) measures to control for perceptual and constructional deficits, such as a copy trial and a matching trial; and
3) a scoring system that reflects as accurately as possible more "nonverbal" aspects of the recall of the material.

Many of the current tests purported to examine nonverbal memory have been extensively criticised and do not meet the above criteria (Baxendale, 1997; Boller & de Renzi, 1967; Heilbronner, 1992; Kapur, 1988; Lee et al., 1989). Although the Visual Reproduction subtest of the Wechsler scales is arguably one of the most widely used tasks to assess non-verbal memory (Lezak, 1995; Tulsky & Ledbetter, 2000), it also does not meet these criteria for an adequate non-verbal memory test.

One of the options available is to produce new tests of "non-verbal memory". However, an alternative to designing a new test is to modify an existing test that is already widely used so that a score more adequately reflects non-verbal memory. This process could be applied to the Visual Reproduction subtest of the WMS-R. Although the WMS-111 included an updated version of Visual Reproduction including recognition, matching and copy trials, there are many potential problems with the subtest. The Visual Reproduction subtest on the WMS-R also has difficulties associated with administration, scoring and interpretation. However, it is used widely as a measure of non-verbal memory and many clinicians have a wealth of experience in scoring and interpreting the reproductions. Modification of the administration and scoring of the revised version, in an attempt to make it more adequately reflect non-verbal memory, would be of particular clinical usefulness in this context. A further advantage in using the WMS-R version is that recent normative data has been published for an Australian population for the ages 18-34 (Carstairs & Shores, 2000; Shores & Carstairs, 2000).

Development of an Alternative Scoring System and an Elaborated Administration of the Visual Reproduction subtest for the Wechsler Memory Scale –Revised

This study will involve the development of an alternative scoring system for the Visual Reproduction subtest of the WMS-R. A number of additional memory testing procedures will also be developed to supplement the standard administration the Visual Reproduction subtest. In doing so, this study has three primary aims.

1) To develop a revised scoring system that eliminates or minimises the weaknesses in the WMS-R scoring system (for the purposes of this discussion called the original scoring system) and has sound psychometric properties.
2) To develop an Index of non-verbal memory function. This Index would potentially provide a diagnostic subscale that identified non-verbal memory dysfunction, mindful that due to the nature of the task it can be verbally encoded at least to some extent;
3) To develop cueing and recognition procedures to supplement the standard administration of the subtest in order to provided additional clinically useful information from the subtest.

Development of the Revised Scoring System

The revised scoring system for the WMS-R Visual Reproduction subtest will be developed to eliminate or minimise the identified weaknesses in the WMS-R scoring system (for the purposes of this paper called the original scoring system). In this way it will be less contaminated by processes other than memory functioning.

The original scoring system on the WMS-R can be criticised on a number of grounds. For example, when applying the original scoring system, not all items are scored independently. This could potentially result in an individual item receiving no score, although the individual had actually produced it. As a result, the final score may not adequately reflect the extent of memory recall. The revised scoring system will be developed so that all items can be scored independently. In this way, the final score derived with revised scoring system will make fewer assumptions about the way information is retained, the score obtained will not be contaminated by failing to producing another item and concerns about a score underestimating memory recall will be reduced.

Some of the individual items on the original scoring system do not include allowances for carelessness in drawing, reduced visuo-motor coordination or poor motor control. Thus, an individual could potentially score poorly, but still recall the main features of a design. In these cases, clinicians can be uncertain as to whether to score the actual product or the intent of the reproduction. The revised scoring system will be developed so that it places less attention on minor imperfections or minor omissions. Generous tolerances in scoring will be included so that a minor imperfection (when the item is drawn as intended but a minor detail is missed, for e.g., including only three dots in each square for design 3), poor motor control or poor drawing ability (for e.g. having gaps between the join of lines and overshoots across lines) is not penalised. This change in focus will be an advantage, because memory recall will be rewarded, rather than drawing precision or skill. This change will also have the potential to improve scoring reliability, because by increasing allowances, there will be less emphasis on precise judgement.

Items on the original scoring system tend to focus on precise angles and distances. This is a potential disadvantage because a person could be classed as having impaired memory due to carelessness in drawing or poor motor control rather than because of "forgetting". Items on the revised scoring system will be developed to include greater tolerances for the measurement of angles and the length of lines. This increase in scoring flexibility will result in memory rather attention to detail being rewarded. Because scorers will be less likely to make a decision on scoring detail versus intent, scoring with the revised system is also likely to result in high levels of scorer agreement.

Another potential limitation of the original scoring system is that there are many instances of partial recall that receive no credit. As a result, the original scoring system may underestimate the quality of memory recall. In clinical practice, participants are often noted to recall partial elements of a design and therefore the revised scoring system will be designed so that partial recall is rewarded. As a result, the revised scoring system will have the potential of being more able to adequately address reproductions of different recall quality.

A potential problem with the original scoring system is that each of the Designs has a different number of items. The WMS-R manual did not present a rationale for this differential weighting. Although it is possible that the designs contain a different number of components and therefore are not equal in difficulty, this is not identified as the reason for the differential weighting. If the designs were weighted equally, there would be a logical basis for prorating scores in the event of disrupted performance on one design. The revised scoring system will be devised so that an equal number of points are allocated to each of the four designs. Therefore, there will be no assumption made about the recall difficulty of any of the four designs. As a result of this change, there will be a greater potential for a database to be developed of differential performance across designs. A method of prorating scores if one design is compromised could also be developed.

The range of potential scores on the WMS-R version of the Visual Reproductions subtest is quite small (0-41). The number of items on the first three designs in particular is limited, with these three collectively contributing only 54% to the total score on the original scoring system. Due to this small range, the discriminatory power of the subtest, particularly at lower and higher levels of performance, may be limited. When developing the revised scoring system the number of items for each design will be increased to 20. This potentially increases the range of scores obtainable on this task, particularly for the first three designs.

In summary, all items on the revised scoring system will be developed so they can be scored independently. More generous tolerances in the measurement of angle degrees and the length of lines will be included to reduce the demands on subjective judgement. This change in focus is likely to make scoring less dependent on precise drawing and return the focus to memory recall. Items will also be included to address partial recall in order to more adequately address designs of different recall quality. Each design will have the same number of items and therefore no assumption will be made about the recall difficulty of any of the four designs. By incorporating these features into the revised scoring system, it will have a number of advantages, including 1) eliminating anomalies in scoring, 2) increasing the range of scores and 3) providing better discrimination between performances.

Although a number of changes will be made to the scoring system, it will not require any changes to the standard administration of the Visual Reproduction subtest. Use of the revised scoring system will therefore be highly compatible with standard administration and will not preclude scoring by the original system and use of the WMS-R normative data. The revised scoring system will also not be designed as a radical departure from the original, rather it is conceptualised as being a similar system, that produces a similar grading of memory whilst circumventing some of the identified limitations of the original scoring system.

Development of an Index of Non-verbal Memory

A second aim of this study is to develop a potential Index of non-verbal memory functioning. Thus, instead of trying to eliminate the contribution of verbal processes on the Visual Reproduction subtest by altering the complexity of the stimuli, an Index will be developed so that those items related to non-verbal memory can be separated from those aspects that are more likely to be verbally encoded. This approach acknowledges that verbal memory processes contribute to performance on this task.

In order to facilitate the development of the non-verbal memory Index, when developing the revised scoring system an emphasis will be placed on identifying items that focus on components that are potentially more difficult to verbally encode. Theoretically, it is thought that the spatial relationships between elements of the designs are likely to be more difficult to verbally encode. Thus, an emphasis will be placed on including as many items as possible to reflect the spatial relationships in the revised scoring system.

The non-verbal memory Index will be derived entirely from the data collected and will not be based on any theoretical model. The primary aim of this Index is to be of assistance when diagnosing non-verbal memory impairment. The end goal will be good sensitivity and specificity in identifying people with non-verbal memory impairment.

Developing an Elaborated Administration of the Visual Reproduction subtest

Based on current notions of memory processing, this study also aims to extend the administration of the Visual Reproduction subtest of the WMS-R to include cuing, recognition and perceptual match procedures to follow the delayed free recall procedure. The cued recall procedure, the recognition measure and the perceptual match task will be formalised to supplement the standard administration of the Visual Reproduction subtest.

The addition of these procedures could potentially enable a greater range of clinical information to be extracted from a commonly used test. For example, the contribution of non-memory deficits such as defective visual perceptual and visuo-constructional skills to poor performance may be illuminated by performance on the perceptual match task. The procedures could also potentially provide diagnostic information that could assist in determining the underlying nature of memory difficulties, such as encoding, storage or retrieval impairments.

The WMS-R version of the Visual Reproduction subtest did not include measures to assess memory storage in a precise manner. Although the WMS-111 included an updated version of the Visual Reproduction subtest that incorporated a number of these additional measures, namely a forced choice recognition task and a perceptual match task, it did not include a cued recall procedure. Given the potential problems with the WMS- 111 material that were outlined previously (Pages 58-59), a decision was made to develop additional materials to supplement the WMS-R version of the Visual Reproduction task, rather than to use the WMS-111 material.

A delayed cued procedure will be added to the standard administration of the Visual Reproduction subtest due to the clinical observation that information previously considered forgotten, is often recalled when a cue is given. Cued recall is a form of testing that uses a particular cue or part of the material previously presented in order to facilitate the recall of information (Watkins & Gardiner, 1982). Cued recall could be considered an intermediate step between free recall, with the only cue to recall being "please tell me" and a recognition task, when the stimulus to be remembered is given with a set of distracters. The value of a cue is that clinical practice suggests that it can be of considerable assistance to some individuals in prompting their recall. If a cue is able to prompt partial or full recall of the material, an individual could be considered as having difficulty with retrieval, although having encoded the material.

A delayed recognition memory task, following the cued recall procedure, will also be developed in order to assess delayed recognition memory. Performance on this task, as compared to delayed free recall will assist the clinician in studying the contribution of retrieval to performance. Thus, if a person were able to recognise a stimulus, even after having no free recall or no cued recall, a difficulty with retrieval of learned information would be inferred. The task developed in this study will attempt to address the problems identified in previous recognition tasks (Wechsler, 1997). It will be designed along similar lines to the materials used by Fastenau (1995) with some modification.

A perceptual matching procedure will also added to the standard administration in line with recent clinical practice (Wechsler, 1997). Visual perception is an integral component of adequate performance on the standard and additional Visual Reproduction procedures. Evaluation of performance on this task could assist in understanding whether perceptual difficulties are confounding performance on the memory task. Poor performance on this task may reflect a difficulty in the perceptual processing of material and therefore performance on the immediate and delayed recall trials is likely to reflect poor perception in addition to memory.

This study will provide information about performance on the additional measures in a clinical and nonclinical population and information about their reliability and validity. It is expected that these measures will provide additional information about memory performance, particularly in a non-clinical population.

Thus, with the development of the revised scoring system and additional measures for the Visual Reproduction subtest of the WMS-R it is hypothesised that

1) The revised scoring system will generate a similar grading of memory to the original scoring system. However, the revised scoring system will have a number of advantages and generate a wider range of scores.

2) The revised scoring system will be psychometrically sound. It will have good internal reliability, as well as strong inter-rater and intra-rater reliability. Reliability measures will be equal or better than for the WMS-R version.
3) Reducing the scoring emphasis on non-memory factors when developing items on the revised scoring system will enhance the consistency of scoring with the revised scoring system.
4) A non-verbal Index of items derived from the revised scoring system will discriminate between clinically meaningful groups;
5) The additional procedures, cued recall, recognition and perceptual matching, will generate extra information about memory functioning.

METHOD

Participants

The study sample comprised 60 adults aged between 50 and 87 years. These participants qualified for either a control group or an experimental group based on the presence or absence of a neurological condition.

Experimental Group

The experimental group was comprised of 30 participants, 28 who were in-patients at the Western Hospital, Melbourne, Victoria, and two who were residing in the community. All had evidence of cerebral dysfunction documented by neurological examination, or neuro-radiological procedures. They represented a wide range of clinical diagnoses including: Left Hemisphere Stroke, Right Hemisphere Stroke, Alzheimer's Disease, Cerebrovascular Dementia, Parkinson's Disease, Epilepsy, Encephalitis, Aneurysm, Hydrocephalus, and Head Injury. Exclusionary criteria included the presence of:

- severe aphasia,
- visual neglect,
- a visual field defect that would interfere with visual discrimination, or
- a cognitive assessment within the previous six weeks where the participant may have been exposed to some of the test materials.

The demographic characteristics of the experimental group are presented in Table 5.

Table 5

Demographic Characteristics of the Experimental Group

Variable	Total group *N*=30	Males *n*=15	Females *n*=15
Age			
Range	54-87	54-87	54-87
M	70.67	69.73	71.60
SD	10.08	10.20	10.22
Years of Education			
Range	7-13	7-13	8-13
M	9.50	9.47	9.53
SD	1.63	1.78	1.55
Employment			
No. of unskilled workers	25	12	13
No. of semi-skilled workers	5	3	2
No. of professional workers	0	0	0

Control Group

Thirty participants, 17 female and 13 male, comprised the control group. They were recruited from non-neurological medical patients at the Western Hospital and from Community Organisations. Exclusionary criteria included a history of:

- neurological disease (e.g., stroke, transient ischaemic attack, tumour, head injury with loss of consciousness greater than 5 minutes, epilepsy, neurosurgery, encephalitis/meningitis, Dementia, Multiple Sclerosis, Parkinson's Disease, Huntington's Disease),
- treatment for a psychiatric condition,
- chronic uncontrolled hypertension,
- any clear history of prolonged, excessive alcohol consumption.

Participants were not excluded from the study due to the presence of stroke risk factors such as hypertension, non-insulin dependent diabetes, heart disease or smoking. The demographic characteristics of the control group are shown in Table 6.

Table 6

Demographic Characteristics of the Control Group

Variable	Total group *N*=30	Males *n*=13	Females *n*=17
Age			
Range	50-85	50-78	53-85
M	68.83	64.08	72.47
SD	10.35	10.10	9.23
Years of education			
Range	8-14	8-11	8-14
M	9.57	9.85	9.35
SD	1.30	.99	1.50
Employment			
No. of unskilled workers	22	7	15
No. of semi-skilled workers	8	6	2
No. of professional workers	0	0	0

Materials

The materials used in this study included:

1) an adaptation of the Verbal Paired Associates subtest of the Wechsler Memory Scale,
2) an elaborated administration of the Visual Reproduction subtest of the Wechsler Memory Scale- Revised (WMS-R),
3) the Logical Memory subtest of the WMS-R,
4) the Faces subtest of the Wechsler Memory Scale- Third Edition (WMS-111),
5) the Block Design and Vocabulary subtests of the Wechsler Adult Intelligence Scale- Revised (WAIS-R) and
6) Part A of the Trail Making Test (Spreen & Strauss, 1991).

These materials are described below. Additional measures were designed as a part of this research project and these are described in more detail in the Design section (page 74).

Visual Reproduction

An elaborated version of the WMS-R Visual Reproduction subtest (VR) subtest was administered. The standard Visual Reproduction subtest consists of four cards with printed visual designs of varying complexity. It has been commonly used as a contrast to verbal memory tasks and the visual presentation mode has also been thought to tap non-verbal memory processes.

In this study, the standard procedure, which involved both immediate and delayed recall trials, was administered according to the procedures outlined in the WMS-R manual. In the immediate recall trial, the participants were requested to draw what they remembered of each of the four designs immediately after it had been presented for 10 seconds (VR1). Approximately 30 minutes later the participants retention of the four designs was assessed (VR11) by asking them to draw what they remembered of the previously presented designs.

The delayed recall trial was supplemented by several additional procedures including:

1) a cued recall procedure involving the presentation of a separate cue for any design omitted during delayed recall;
2) a six alternative choice recognition measure for each of the four design cards; and
3) a matching procedure requiring the participant to match the correct VR design to its identical match in an array of six.

The development and rationale for these additional procedures is discussed more fully in the Design section (See page 74).

The participants recall for the immediate and delayed conditions and the cued recall task were scored according to the scoring system specified in the WMS-R manual. They were also scored according to the revised scoring system devised as a part of this study (See page 84).

Logical Memory

The standard administration of the Logical Memory subtest of the WMS-R was used in this study. This administration involved both immediate (LM1) and delayed (LM2) recall trials.

In LM1, two stories were read to the participant. Immediately after the presentation of each the participant was asked to retell what they could recall of this material. In LM2, the participant was asked to recall the two stories after a 30-minute delay.

This subtest was administered according to the procedures outlined in the WMS-R manual. However, minor changes were made to the names of places in the stories to make them more geographically appropriate and to the monetary value in Story A to make it more consistent with current values. There was also a minor modification to reduce the word length of the first sentence in each story. The rational for these changes is outlined in the Design section (page 81).

Both stories were scored for verbatim recall as outlined in the WMS-R manual.

Verbal Paired Associates

This study incorporated an elaborated administration of the original WMS Paired Associates Learning subtest. In the original procedure, ten word pairs, six "familiar" pairs and four "novel" pairs, were read to the examinee three times, with a recall trial following each reading. Each recall trial involved the participant being read the first word of each pair and being required to supply the second word (VPA1). Administration of the immediate recall trial followed the guidelines outlined in the WMS Manual.

A 30 minute delayed recall trial was added to the original procedure as was standard in the WMS-R version. The participant was again provided with the first word of the pair and asked to supply the second word (VPA2). The directions for the VPA task in the WMS-R manual were used for this trial because it provided instructions for delayed recall. The total number of "familiar" and "novel" pairs recalled over the three immediate recall trials and in the delayed recall trial (VPA2) were recorded separately.

Faces

The Faces subtest of the WMS-111 was administered in its original form as per the Manual and there were no additional testing procedures included. However, in addition to the standard scoring, extra data was collected in accordance with signal detection theory, that is, false positive, true positives, false negatives and true negatives.

The Faces subtest evaluated immediate and delayed recognition memory for faces. The faces were of mixed gender, age and ethnicity. In the immediate recognition trial, 24 target faces were individually presented to the participant, at the rate of one per two seconds. The participant was then asked to identify the target faces from another series of faces consisting of the 24 target faces and 24 distracters. The participant was required to give a "Yes" response if they believed that the face was one of the ones they were asked to remember and a "No" response if it was not. After a delay of 25-35 minutes, the participant was again asked to identify the 24 target faces amongst a new set of 24 distracters.

Block Design

The Block Design subtest is a constructional task in which the participant was required to use blocks to construct a pattern that matched one made by the examiner and/or printed in smaller scale on a card. It was administered and scored according to the procedures outlined in the WAIS-R manual.

Vocabulary

The Vocabulary subtest of the WAIS-R was administered in this study. This task involved the oral and visual presentation of a series of words that the participant was required to define. This test was administered and scored according to the procedures outlined in the WMS-R manual. The WAIS-R version of this test (and the Block Design subtest) was chosen instead of the WAIS-111 version because previous research with the WMS-R has included WAIS-R subtests.

Trail Making Test

The Trail Making Test, Part A, required the participants to draw lines to connect consecutively numbered circles in a scattered array. It was administered according to the instructions provided by Spreen and Strauss (1991) with the exception that errors were not corrected. It was scored for the time taken to complete as well as number of errors. Additionally, an alternative form was created (See Appendix A). This form was a mirror reversal of the original material so that the spatial relationship between numbers was preserved. As the Trail Making Test was administered on four separate occasions, the alternative form was developed in order to reduce the potential for practice effects.

Design

Subtests Administered

Visual Reproduction: Additional Procedures

In addition to the standard administration of the Visual Reproduction subtest, a cued recall procedure, a recognition measure and a perceptual match task were designed and administered in this study. These procedures were developed with the aim of providing additional clinical information regarding the contribution of non-memory or cognitive factors to performance on a memory task and also to provide information about the underlying nature of memory difficulties. Given the wide clinical usage of the Visual Reproduction subtest, these procedures could potentially result in the extraction of a greater range of clinical information from a commonly used test.

The WMS-111 provided an updated version of the Visual Reproduction subtest that incorporated a number of these additional measures, namely a forced choice recognition task and a perceptual match task. However, some potential problems with the new material lead to the conservative decision to use the WMS-R version of the Visual Reproduction task. For example, the addition of a simpler design that was highly similar to the second design, could lead to confusion in the participant's recall. Three of the designs also included flags, making it difficult to differentiate between them, particularly on delayed recall. There were also a number of potential problems with the content and format of the recognition task. These included:

1) The task was extremely lengthy and time consuming.
2) The two designs in items four and five were presented separately which could be confusing for participants.
3) The distracters were very similar to the target designs which may have resulted in partial recognition of something in a similar distracter leading to an incorrect yes response.
4) Participants may have become so confused due to the similarity of the designs in the recognition task that they may have said yes to a distracter

design because it was similar to another distracter design presented previously during the recognition task.

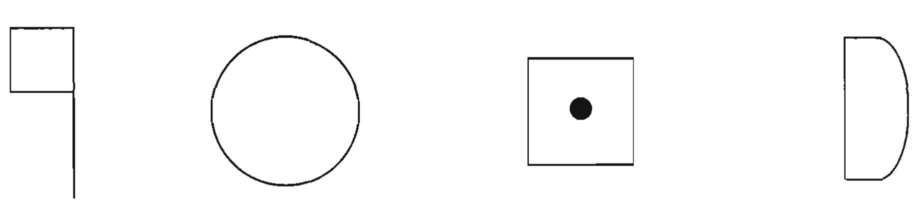
In regard to the format of the WMS-111 recognition task, recognition was tested via a YES/NO response and therefore guessing could confound recognition. Given these potential problems, a new procedure was developed in order to assess recognition memory. The WMS- 111 version also did not include a cued recall procedure, which could provide a useful intermediate step between the assessment of delayed free recall and delayed recognition.

Cued procedure.

A cued recall procedure was used in this study due to the clinical observation that participants will often recall information previously considered "forgotten" when given a cue. The visual cues used in this study were based on items used in clinical practice (P. Dowling, Personal Communication, 1997). Each cue was developed to reflect a partial feature of the total design.

In the cued recall procedure, the participant was provided with a visual cue to assist in the recall of any design that they did not recall freely after a delay. A cue was only given when the design was completely omitted during delayed recall.

In this condition the examiner drew the cue for each omitted design whilst asking the participant "Does this help you remember any of the designs?" The visual cue for each design is shown in Figure 1.



Design 1 Design 2 Design 3 Design 4

Figure 1: Cues used for each of the Visual Reproduction Designs

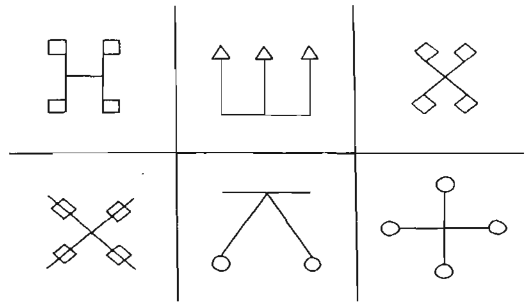
The cueing procedure preserved the recall and drawing component of the original VR procedure allowing it to be scored according to same procedure as VR1 and VR11.

Recognition procedure.

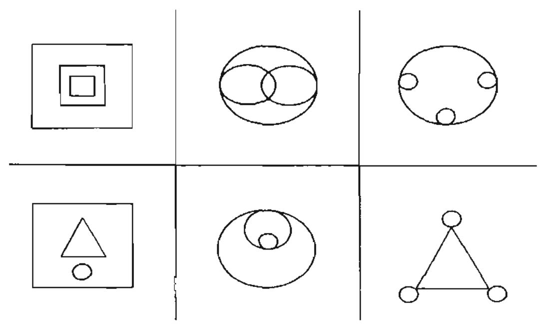
A recognition procedure was developed in order to assess delayed long-term recognition memory and to assist the clinician in studying the contribution of retrieval to memory performance. Although other researchers have reported the use of a recognition procedure for the Visual Reproduction subtest (Fastenau, 1996; Milberg et al.,1986), the actual materials used in this study were an extension of material used in clinical practice (P. Dowling, Personal Communication, 1997). This clinical material was expanded to include a six choice rather than a four choice recognition format.

In this condition, each participant was given the opportunity to identify the correct Visual Reproduction design from a 2 x 3 array of six multiple-choice items that comprised the correct design and five distracters. The correct item was placed in a different position for each of the Designs. No correct item was placed in the first or last position in the array because the natural bias in guessing tends to favour those positions.

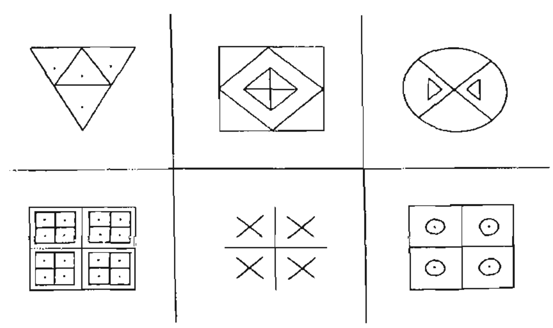
The procedure for assessing recognition followed the cued recall procedure. As is shown in Figure 2, there was one recognition trial for each of the four Visual Reproduction cards. The participant was instructed: " Now I am going to show you some more designs. Are any of these one of the designs that I asked you to remember". A recognition format was shown for each of the four designs, even if the participant had shown partial or total recall of that design during the free recall or cued recall procedure. Scoring was the number of designs correctly identified out of a total of four.



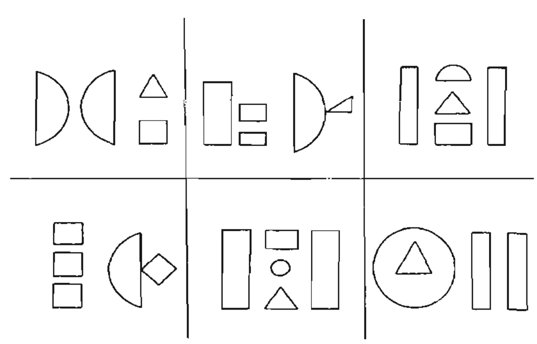
Design 1



Design 2



Design 3



Design 4

Figure 2: Multiple-choice Stimuli for the Visual Reproduction Subtest

Perceptual match procedure.

A perceptual match task was added to Visual Reproduction subtest in order to evaluate if perceptual difficulties were confounding performance on the memory task. This task was based on the Matching procedure outlined in the WMS-111 (Wechsler, 1997).

In this study, the participant was required to match the correct Visual Reproduction design to its identical design in an array of six designs. The stimuli used in this task for Cards A, C and D were identical to those used in the matching task for Visual Reproduction on the WMS-111. The stimulus used for Card B was designed as a part of this study and is illustrated in Figure 3.

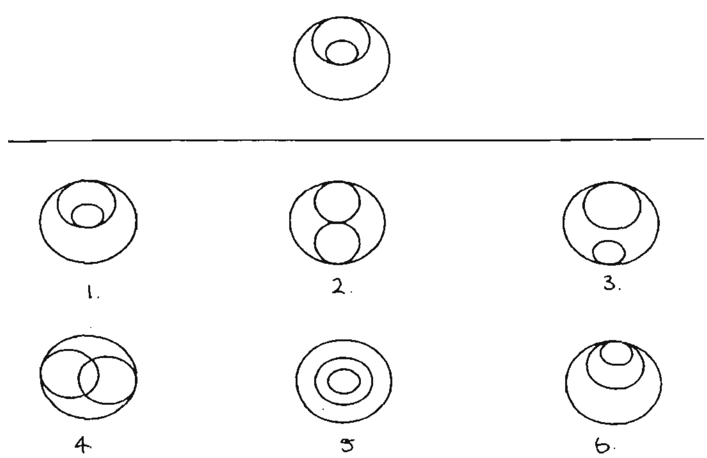


Figure 3: Perceptual Match stimuli for Design 2 of the Visual Reproduction Subtest

As is illustrated in Figure 3, the participant was shown each of the original designs separately on the top of a sheet containing six variations. They were then asked: "Which one of these (pointing at the array) looks like this picture exactly (pointing at the stimulus)". The score was a total correct out of five.

Logical Memory

The Logical Memory subtest was included in this study as a measure of episodic verbal memory. Verbal memory measures were included in this study in order to investigate their relationship with non-verbal memory tasks and to help establish the construct validity of the Visual Reproduction subtest.

The standard administration of the Logical Memory subtest of the WMS-R was included in this study due to its well-established criterion validity as a measure of verbal memory and its good reliability (Bornstein et al., 1988; Larrabee & Curtis, 1995; Naugle et al., 1993). The WMS-R version of the Logical Memory was included because normative data was available for this subtest on the same group as the normative data for the Visual Reproduction subtest and because there were format problems with the WMS-111 version of the Logical Memory Subtest.

When administering the WMS-R version of the LM subtest, minor adjustments were made to the wording of the two stories. In story one, the long first sentence in each story was divided into two sentences. However, minimal changes were made, the original word order was preserved and no new ideas or specific content was introduced. The rationale for this change was that the length of the sentences, 36 words for Story A and 25 words for Story B were well beyond the normal sentence span as published in Spreen & Strauss (1991). Thus, many participants could fail to register essential elements from the sentences, because they are beyond their normal sentence span. In addition, American place names were replaced with Australian regional place names to make the stories more geographically relevant. The monetary value in the first story was increased to reflect the increased cost of living in current times. The changes in the two stories are illustrated in Table 7.

These changes did not affect the application of the scoring criteria and both stories were still scored for accuracy according to the Wechsler (1987) scoring criteria.

Table 7

Changes to the Original Stories of the Logical Memory Subtest

	Original version	Revised version
Story 1	Anna Thompson of South Boston, employed as a cook in a school cafeteria, reported at the City Hall station that she had been held up on State St the night before and robbed of $ 56 dollars. She...(no further changes after this point).	Anna Thompson of South Melbourne was employed as a cook in a school canteen. She reported at the police station that she had been held up on Market St the night before and robbed of $120 dollars. She...(no further changes after this point).
Story 2	Robert Miller was driving a ten ton truck down a highway at night in the Mississippi Delta carrying eggs to Nashville when his axle broke. His...(no further changes after this point).	Robert Miller was driving a ten ton truck down a highway at night in the Murray Valley. He was carrying eggs to Newcastle when his axle broke. His...(no further changes after this point).

Verbal Paired Associates

The Verbal Paired Associates (VPA) task was included in this study because of its validity as a measure of episodic verbal memory (Frisk & Milner, 1990; Larrabee & Curtis, 1995; Rausch & Babb, 1993; Saling et al., 1993). In this way it was considered a good task to compare with non-verbal memory tasks when establishing construct validity.

The WMS-R version of the VPA task included eight word pairs, four of which were familiar associations, four of which were novel associations. A potential problem with this version was that eight pairs was close to the average persons immediate span of seven and therefore the actual capacity of the task to measure learning may be reduced. Additionally, no normative data such as percentiles or standard scores was provided in the WMS-R manual and a score could only be obtained for verbal memory if the other verbal subtests on the WMS-R were given. Thus, the VPA task could not be interpreted separately from the rest of the WMS-R scale.

The WMS-111 version of the VPA task included both immediate and delayed recall trials and a delayed recognition procedure. However, there were also some potential problems with this version. Firstly, it only contained unrelated or "novel" word pairs. Although there are valid reasons for only including "novel" word pairs, the clinical information about old and new learning that could be obtained from a comparison between "familiar' pairs and "novel" pairs could be considered important, particularly in a clinical population. The use of familiar pairs potentially enables people to have some correct performance and the task is potentially less challenging.

Mindful of the potential problems with the WMS-R and WMS-111 versions, an elaborated version of the WMS Verbal Paired Associates task was chosen for this study. The WMS version included 10 word pairs as compared to eight word pairs in the WMS-R and both "familiar" and "novel" associations. Normative data has also been published for an Australian population (des Rosiers & Ivison, 1986). In line with current theories of memory, a 30 minute delayed recall trial was administered in addition to the standard three learning trials.

Faces

There is substantial evidence that memory tasks using faces are better stimuli to use when attempting to measure non-verbal memory processes because of the difficulty in verbally encoding this type of material (Milner, 1968; Morris et al, 1995b; Naugle et al, 1994; Warrington, 1984). Despite this, there are few tests of facial memory and many omnibus tests do not include a facial memory task. Of the available tests of facial memory, the WMS-111 Faces subtest was used in this study. The WMS-111 Faces subtest was chosen because the standardisation sample of the WMS-111 provided some information about its relationship with other non-verbal and verbal memory tasks. In particular, the WMS-111 manual provided some information about its relationship with a non-verbal memory test that used design recall, that is, the Visual Reproduction subtest.

Block Design

The Block Design subtest of the WAIS-R was included in this study as an estimate of general intellectual ability. It is a highly reliable subtest and correlates highly with general mental ability (Kaufman, 1990). It was also included as a marker of visuo-constructional and visuo-perceptual cognitive ability and it would be expected that it would share greater common variance with non-verbal memory tasks than verbal memory tasks.

Vocabulary

The Vocabulary subtest of the WAIS-R was included in this study as an estimate of general mental ability. Although Block Design can provide an estimate of general ability in normal samples, it is sensitive to brain damage. Therefore Vocabulary was included as an alternative estimate of general ability.

The Vocabulary subtest was considered a good measure of verbal intelligence and performance on this subtest tends to be less disrupted by most acquired neuropsychological conditions. However, it is related to educational experience and occupational achievement (Kaufman, 1990). Thus, by administering both the Vocabulary and Block Design subtests it was considered that a good estimate of IQ could be made in the study sample.

This subtest was also included as a marker of verbal cognitive abilities and it would be expected that it would share greater common variance with verbal memory tasks rather than non-verbal memory tasks.

Trail Making Test

Part A of the Trail Making Test was administered as a measure of basic attention. It was the first task administered and gross failure (i.e. greater than 5 errors or more than 3 minutes completion time) resulted in exclusion from the study. An alternative version of this task was also developed to allow repeat testing. This alternative version was a mirror image of the original and is seen in Appendix A.

The standard or alternative version of this task was administered at the beginning and end of the two testing sessions to confirm that the participant's attention had not deteriorated to any major extent over the course of the session. If the time taken to complete the task was substantially longer or the number of errors were much higher on the second administration, then the reliability of the assessment was considered questionable and the participant was to be excluded from the study. In no case did this occur.

Development of the Revised Scoring System for the Visual Reproduction Subtest

This study involved developing a revised scoring system for the Visual Reproduction subtest of the WMS-R. Although problems have been identified with the subtest format and the scoring criteria, a change in the scoring system to better reflect non-verbal memory processes may be preferable to devising and validating a new test. Given that the Visual Reproduction subtest of the WMS-R continues to be widely used, it was considered that refinement of the scoring system would be beneficial in order to provide more clinically useful information.

A number of problems and limitations have been identified in the WMS-R scoring system. Scoring tended to focus on precise measurement of angles and lengths, participants could potentially be penalised for carelessness in drawing because items focus on attention to detail and there was a lack of reward for partial recall.

The first stage involved developing the revised scoring system on the basis of a number of principals which included:

1. identifying the common types of errors made from a sample of clinical protocols and using these to guide the development of the criteria;
2. identifying the features that were correct across most clinical protocols and limiting the points assigned to these;
3. identifying aspects of the designs that were particularly difficult to verbally encode, for example, the spatial relationship between elements of the designs, and developing items that addressed these features;
4. item independence. Due to concerns over dependence in scoring items on the original scoring system, the revised scoring system was designed so that each item was scored independently. In this way, the final score made no assumption about

the way information was retained and reduced concerns about underestimating memory recall.

5. placing less emphasis on attention to detail and instead allowing greater tolerance in the measurement of angle degrees and the length of lines. By increasing the tolerances allowable, scoring would be less dependent on precise drawing and instead be focused on memory recall;
6. including lenient criteria that addressed poorer examples of recall and including strict criteria to increase the sensitivity of the test and the range of scores;
7. minimising penalties for poor drawing or an occasional lapse (e.g. forgetting one dot) and allowing gaps between lines and overshoots across lines. This change in focus from the original scoring system would place the emphasis on memory recall, rather than precision in drawing. This should also improve the reliability of scoring because clinicians will not have to decide whether to score the product or the intent;
8. examining ways of giving credit for incomplete or partial recall.
9. specifying generous tolerances for most criteria so that demands on subjective judgement from the scorer would be reduced;
10. including criteria that addressed all aspects of the design;
11. where there were multiple figures (Design 4) each half of the design was given approximately equal rating;
12. whilst criteria were developed so that extra material did not preclude people scoring points on the parts of the design that were correct, one item for each of the four designs was assigned as a penalty for the inclusion of any extra material.

As well as identifying possible items, the number of items to be generated for each design was addressed. The number of items for each of the designs in the WMS, WMS-R and WMS-111 varied and was not equal across all designs. For example, the Wechsler (1987) scoring system contained 7 items for designs A and B, 9 items for design C and 18 items for design D giving a total of 41. No rationale was given in the WMS-R for the differential loading given to each design. Given that each design did not contribute equally to the total, if one design was compromised in some way, no compensation could be made. Additionally, the small number of items for some of the designs in the WMS-R scoring system reduced the range of possible scores and this may have reduced the scope of the scoring system to discriminate between performances.

These considerations were taken into account when deciding how many items to develop for each design. The revised scoring system was developed to include 20 items for each of the four designs. Twenty items was considered adequate to allow enough coverage to discriminate between designs without being excessively lengthy or over-inclusive. An equal number was allocated to each design so that if one item was compromised, for example, by interruption, a prorata score could be developed.

The revised scoring system was developed to increase the sensitivity of measuring non-verbal memory as represented by recall of designs. Rather than a radical departure from the original scoring criteria, the revised scoring system was designed to reflect a similar grading of memory functioning, so that weaker memory functioning was accurately reflected by both scoring systems. Thus, it was expected that the two sets of criteria would generate or produce similar ratings of the quality of recall and therefore have substantial correlation.

Procedure

Development of the Revised Scoring System

Stage one in the development of the revised scoring system for the Visual Reproduction subtest involved reviewing clinical material from a variety of cases collected prior to this research. These clinical protocols included a variety of clinical conditions, a range of ages and covered the range of performance seen on this task. Individual items were identified according to the principals outlined in the Design.

In stage two, clinical material was used to evaluate whether the revised scoring system was able to discriminate between individual cases and whether designs of a similar quality scored similarly. Scoring of a number of designs of different quality was conducted in order to see if the revised scoring system resulted in a range of scores. The clarity of the revised scoring system was also investigated by seeing if two examiners interpreted the items on the revised system similarly.

Stage three involved an initial comparison between the original and revised scoring systems. Forty-four clinical examples of reproductions were scored on the original and revised scoring systems. The relationship between the original and revised scoring systems was then examined by correlation analysis. It was expected that the two scoring systems would share a high correlation as many of the items were similar and because both were measures of memory performance.

Stage four involved reviewing and refining the revised scoring system. Due to examples of difficulty in interpretation of specific criteria, further refinement of individual items was conducted. This refinement primarily focused on rewording the items to make their meaning more explicit and more easily interpreted by scorers.

Stage five involved a comparison between the two sets of scoring criteria after the Stage four revision. A set of 50 clinical cases was selected (data collected separately from this study) in order to compare the two scoring systems. For each design, 50 immediate and 50 delayed reproductions were scored on the revised and original scoring systems. A correlation analysis was conducted to evaluate the similarity of the two systems.

In stage six the relationship between the two scoring systems was compared in the study sample. The reliability of the two scoring systems was investigated. In order to establish the reliability of scoring on both systems, intra-rater and inter-rater reliability were examined using the study sample. Measures of consistency that were calculated to evaluate these forms of reliability included correlation analysis, the percentage of agreement in the total score and the level of agreement on each item. These measures were calculated between the scores collected on two separate occasions by the one scorer and between the scores obtained by two separate raters. A measure of internal consistency, Cronbach's Alpha, was also calculated.

Stage seven involved investigating the validity of the Visual Reproduction subtest, as scored by the original and revised scoring systems, as a measure of non-verbal memory. Concurrent validity was examined by looking at the relationship between non-verbal and verbal memory measures via correlation analysis. Construct validity was examined by looking at the difference between the clinical groups on the Visual Reproduction subtest.

The final stage of the development of the revised scoring system involved examining the differences in performance in a clinical population with a view to identifying particular items that could be of use in discriminating between verbal and non-verbal memory processing. This non-verbal memory subscale was constructed by comparing the performance of two clinical populations, one with right hemisphere lateralised damage, and one with left hemisphere lateralised damage, on each individual item in the two scoring systems.

Development of Additional Measures

This study also involved the development of a cueing procedure and a multiple choice format for the Visual Reproduction subtest. This was conducted by reviewing clinical material (P. Dowling, Personal Communication, 1997). Possible items were identified and this material was modified and expanded upon and a procedure for their administration was systematised.

The clinical usefulness of the material was examined by collecting the number or participants who benefited from this material and the score benefit gained. Reliability was examined by the Cronbach's Alpha internal consistency measure. Criterion related validity was investigated by examining the correlation with other measures and construct validity was examined by looking at the difference between two clinical groups, an experimental and a control group, on these tasks.

The Use of the Expanded Assessment Format with a Clinical and Non-Clinical Sample

This project involved administering the test protocol to participants in order to investigate the reliability and validity of the revised scoring system. Once participants were identified, either through reviewing medical files, or by recommendation from a community-based organisation, they were approached and invited to participate in the study. This process involved an explanation of the study and the provision of a plain language statement (Appendix B). If the participant was willing to participate they were asked to give their informed consent and sign a consent form (Appendix C). If the participant were available at this time, the assessment was commenced. If not, an appointment was made for a later time.

Before the testing session began, participants were asked questions about their medical history, years of education, level of education reached and main occupation. Following this they were administered the test protocol.

Each participant was individually tested across two sessions of approximately 50 minutes, no greater than seven days apart. The test protocol was administered in a standard way and is detailed in Table 8. At the completion of the second session, participants were thanked for their participation. No specific feedback was given about levels of performance, but if participants enquired, general reassurance was given.

Table 8

Order of Administration of Tests

Session A
Trail Making Test Part A
Visual Reproduction Immediate Recall
Logical Memory Immediate Recall
Vocabulary
Visual Reproduction Delayed Recall
Visual Reproduction Cued Recall
Visual Reproduction Multiple Choice Recognition
Visual Reproduction Perceptual Match
Trail Making Test Part A
Session B
Trail Making Test Part A
Faces Immediate Recall
Verbal Paired Associates Immediate Recall
Block Design
Faces Delayed Recall
Verbal Paired Associates Delayed Recall
Trail Making Test Part A

RESULTS

The Revised Scoring System

The revised scoring system was developed on the basis of the principals outlined in the design. The resulting scoring system for each of the four designs is illustrated in Table 9 for Design One, Table 10 for Design Two, Table 11 for Design Three and Table 12 for Design Four. A general scoring rule was also specified, this being that all items were to be scored independently. Thus, failure on any item did not imply failure on other items.

Data Analysis

All raw data was analysed using SPSS 8.0 for Windows. All variables were inspected for skewness and kurtosis. Skewness and Kurtosis values revealed that many variables were negatively skewed or were flat in distribution (Appendix D). Conformity to parametric assumptions was formally examined with Kolmogrov-Smirnov normality tests. In many cases the critical value of the Kolmogrov-Smirnov test exceeded .05, indicating departure from normality (Appendix E). 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. In addition to this conservative approach, variables that were included in the correlation analyses were checked for linearity of relationship. No curvilinear relationships were identified.

Table 9

Scoring items for Design One on the Revised Scoring System

Item	Description
1	There are at least two continuous lines or 4 lines emanating from a central point or figure. If there is only one line or more than 4 lines emanating from a central point score this item 0, however proceed with scoring other items.
2	There are only two lines present and these two lines intersect. If there are more or less than two lines score 0.
3	Two lines intersect in the middle 1/3 of each. If there is a radial spokes design, all spokes are the same size (50% tolerance between the shortest and longest line). A minor gap at or near intersection point is acceptable if there is no change in direction.
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 length of the longest line (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 and if one is horizontal the other is not vertical. If the lines do not intersect, score 0.
7	At least three geometric figures are present.
8	All figures that are present have the same number of sides (or all are circular). Figures an share a side with the staff/lines but they can not share a border with an external figure (e.g. a bordering square).
9	All figures are identical in shape and size (90% tolerance) to each other.
10	All figures are between 30-50% 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 point to compare with the length of the line. If there are no lines, score 0.

Table 9 Continued on Next Page

11	At least ¾ of the figures are squares/flags, that is, they have four sides. Gaps between the line and flag/square are acceptable if not greater than 25% the length of the side (where the gap is present) of the figure.
12	Exactly 4 discrete figures are present. The figures do not share a border with an external square.
13	More than ½ of the figures touch any line forming a staff. Figures can overlap the line.
14	At least two figures are correctly positioned near the end of the appropriate line (forming a staff). Figure does not have to touch or overlap the line.
15	All four figures are correctly positioned near the end of the appropriate line (forming a staff). Figure does not have to touch or overlap the line.
16	All four figures touch one line at the endpoint of the line. Minor overlap or gaps (< 10% of the length of the longest side of the figure) are acceptable. If the figure/s is a circle, overshoot or gap must be < 10% of the diameter of the circle.
17	The side of the figure is contiguous with the line, that is, the figure shares a side with the line. A minor gap between the line forming the figure and the line forming the staff (<10% length of the figure) is acceptable.
18	Two figures face inwards (if rotation of the lines is <90 degrees assume direction that would maximise the score). If the figures are not contiguous with the line score 0 unless 100% of the figure is on the correct side of the adjacent line.
19	Four figures face inwards. If the figures are not contiguous with the line score 0 unless 100% of the figure is on the correct side of the adjacent line.
20	No extra elements. Minor overshoots of lines should not be penalised.

Table 10

Scoring items for Design Two on the Revised Scoring System

Item	
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 2 geometric figures are present, 1 mostly inside the other. Figures can share a border.
4	The figures form a clear size gradient, that is they are not of equal size. If there is more than three figures, take the largest one to be the large figure, the smallest one to be the small figure and choose medium figure to maximise the score. If there are only two figures, they must form a clear size gradient. A dot is not a figure.
	For items 5- 14: If there is only 2 figures, score spatial questions to maximise the score.
5	Large figure mostly encloses at least one smaller figure.
6	Large figure mostly encloses two smaller figures.
7	Small figure is largely enclosed by a medium figure.
8	Medium figure is located to towards the top of the large figure and away from the bottom. The gap from the bottom of the large figure to the bottom of the medium figure should be at least three times the gap between the top of both figures.
9	Top of the medium figure touches the top of the large figure. Minor overlap, or gap, (< 10 % the diameter) between the two figures is acceptable.
10	Small figure is located close to 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 of the small figure and the bottom of the medium figure.

Table 10 Continued on Next Page

11	The bottom of the small figure touches the bottom of the medium figure. Minor overlap or gap, (< 10% the diameter) is acceptable.
12	Medium figure is about 1/3 the diameter of the vertical axis of the large figure. Diameter should be in the 25-50% range.
13	Small figure is about 1/3 the diameter of the vertical axis of the medium figure. Diameter should be in the 25-50% range.
14	Areas enclosed by each figure are in correct relative proportion. The area of the small figure is about 20-25% of the area of the medium figure and the area of the medium figure is about 20-25% of the area of the large figure.
15	The figures are symmetrically placed about 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 the axis.
16	The spatial relationship between the three figures is preserved, even if the relationship is inverted. If there are only 2 circles, score 0.
17	All shapes are primarily closed circular figures (can be ovals). The border of the smaller figure does not share common contact (the same border) that is greater than 20% of the circumference of the medium figure. The medium figure does not share common contact (the same border) that is greater than 20% of the circumference the large figure.
18	At least 2 of the figures are discrete circles rather than ovals, that is, separate circles in their own right The smaller diameter of a circle is at least 90% of the larger diameter. The border of a circle does not share common contact (the same border) that is greater than 20% of the circumference of another circle.
19	All figures are discrete circles rather than ovals or another geometric shape The smaller diameter of a circle is at least 90% of the larger diameter. The border of a circle does not share common contact (the same border) that is greater than 20% of the circumference of another circle.
20	No extra elements, except minor line continuations.

Table 11

Scoring items for Design Three on the Revised Scoring System

Item	
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. Treat as a large square if there are small gaps in the lines. If there is more than one large outside square, treat the outermost square as the large figure. If there is any doubt, score to maximise scoring.
2	At least one large four-sided figure is present and is approximately square. The four-sided figure may be rectangular as long as the shorter two sides are not 50% smaller in length than the longer two sides.
3	The four sides of the figure in number two are reasonably equal in length. The longest side is no more than 25% longer than the shortest side.
4	The four-sided figure is exactly a square. The smallest line is at least 90% the length of the largest line and every angle is 90°. Gaps/overlaps are acceptable as long as they are less than 10% of the length of the line. All angles are between 85° and 95°.
5	A vertical division divides the large figure. A double lined vertical division is acceptable. Division can be contiguous with the internal squares. Gaps in the join between the division and external figure are acceptable as long as 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. Division can be contiguous with internal squares. Gaps in the join between the division and external figure are acceptable as long as 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, that is, they touch and cross each other.

Table 11 Continues on Next Page

8	Two to four smaller figures are present with or without a major figure bordering them. Each of the smaller figure shares no more than two sides with another smaller figure, the internal horizontal or vertical division or the external square. If there are more than 4 smaller figures score 0.
9	Each quadrant of the large figure has only one smaller figure (of any description). Quadrants do not need to be symmetrical. No more than two figures share a line with each other, that is, at least two of the figures are discrete. OR If the large figure is absent, the smaller figures form a 2 x 2 matrix. The quadrants are defined by the large figure and/or by the bisecting lines. No more than two figures share a line with each other, that is, at least two of the figures are discrete.
10	Smaller figures are distinct shapes (even if less than or greater than four). Figures do not overlap each other, the sides of the square, the vertical/horizontal lines, or any additional lines drawn and they do not share a common border (touching or partial overlap is permitted).
11	Each of the smaller figures is divided into four parts, or there are four shapes in each quadrant in a 2 x 2 matrix. The quadrant being divided into four is not acceptable.
12	Each smaller figure is divided into four, or each quadrant is divided into four by a horizontal and a vertical line. Can be double vertical and horizontal lines.
13	The smaller figures (as per the clients reproduction) are similar in configuration or similar in size (90% tolerance).
14	At least three of the smaller figures are in the correct proportion to the larger figure (as per the original). If there is no large square score 0.

Table 11 Continues on Next Page

15	The smaller figures (in terms of the clients reproduction) have four sides and are separate from each other, i.e. distinct squares, the horizontal and vertical division and the external square. There is no overlap between sides.
16	Some number of dots/circles are present in at least 75% of the internal segments of the smaller figure; in at least 75% of each smaller figure if there are no segments; in at least 75% of the smaller figures that the client produces; or in at least 75% of each quadrant.
17	Each quadrant of the large figure has only four dots/circles in a square array (Divisions may or may not be present). Score zero if there is no large figure.
18	The dot or small open circle occupies <10% area of space in segment of internal figure, in each figure if no segments or in each quadrant if no figures.
19	All four smaller figures are placed symmetrically. If the large square is divided by a vertical and horizontal line, the squares are placed with approximately equal border spacing as per the original design. All the borders are <20% of the length of the quadrant. **OR** If there are no vertical and horizontal lines, the smaller figures are placed with equal border spacing (larger spacing between vertical and horizontal acceptable) **OR** If there is no external square and the smaller figures are in 2x2 matrix, there is equal spacing the between the smaller figures in the horizontal and vertical plane.
20	No extra lines, dots or figures. Minor overshoots of lines should not be penalised

Table 12

Scoring items for Design Four on the Revised Scoring System

Item	
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 only 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 than the parallel side.
3	One or more three to six sided figures is adjacent to the large rectangle (sharing a border is acceptable). If there is no large rectangle, there are one or more 3-6 sided figures present.
4	The smaller figure/s in Number three are separate from each other and from the major figure (minor touching or overlap is acceptable).
5	Bases of all the figures are of similar length (90% differential). If there is only one figure score zero.
6	The tall rectangle is clearly above the height of the adjacent figure/s (do not need to be four sided) beside it (at least 10% of its height).
7	The base of the tall rectangle and the lower of the adjacent figure/s (do not need to be four sided) are level (10% of it height).
8	There are two four sided figures positioned on top of each other and to the right of the large rectangle or the two figures are positioned on top of each other if no large rectangle. The rectangles widths are greater than their heights.
9	Of the two figures in Item 8, one is clearly larger and is positioned above the smaller The smaller is no more than 70% the height of the larger at any point.
10	A large figure with a curved surface, or a curved line, is present.
11	This large figure is a discrete semicircle only (irrespective of orientation).

Table 12 Continues on Next Page

12	The curved portion of the semicircle is facing the right.
13	Semicircle of correct proportion The radius of the semicircle is 1/2 vertical height (40-60% differential of vertical height is allowable).
14	The figure in 10 is located to the right of the figures mentioned from 1-9, some other shape or even a line. Score zero if there are no other figures to the left.
15	A smaller figure is located near the figure in Item 10 or if Item 10 is not present, a smaller figure is placed to the far right of any other geometric shapes This smaller figure being inside the figure in 10 is acceptable. A line receives credit for this item but would not score for 16-19.
16	The smaller figure is separate from any other figure. Smaller figure is not inside another figure and does not share a side with another figure. The smaller figure can touch/overshoot or can be in close proximity (i.e. gap OK) the figure in Item 10 (10% diameter tolerance).
17	The smaller figure is located to the right of the figure in 10, or to the right of some rectangle if no figure in Item 10.
18	The smaller figure is located at or near to the centre of the right border of the figure in Item 10. If item 10 is not present, the smaller figure is located away from the base of the rectangular figures and not above the upper edge of the rectangle. The smaller figure must be located within + or - 30 degrees of the midpoint of the arc of the figure in Item 10 (the figure can be inside another).
19	The smaller figure is a discrete triangle in shape, i.e., it has three discrete sides separate from the semicircle
20	No extra elements. Minor overshoots of lines should not be penalised.

Initial Comparison between the Original and Revised Scoring Systems

After the revised scoring system was developed, 44 clinical examples of each design were collected from a variety of sources. These examples included immediate and delayed reproductions and there was considerable variation in the quality of recall. In order to examine the degree of association between the revised scoring system and the original scoring system, the visual reproductions were scored on both scoring systems and a correlation analysis was conducted. Five variables were examined, namely the score for each of the four designs and the total score.

Since the sample was one of convenience, it could not be assumed that the scores obtained would approximate a normal distribution. In fact, Kolmogrov-Smirnov normality tests (Appendix E) revealed that there was some violation of the assumptions underlying parametric tests and therefore Spearman's correlation coefficients were calculated. The analysis for each of the five variables was conducted in two ways; 1) including all protocols and 2) excluding protocols that scored zero on both the original scoring system and the revised scoring system. It was considered that excluding designs that scored zero on both scoring systems would be a more conservative approach because including those designs that scored zero on both systems had the potential to inflate the correlation and generate an overestimate of the association between the two scoring systems.

When interpreting the correlation coefficients in this study, correlations at or below .10 were regarded as weak in magnitude, correlations around .30 as moderate in magnitude, and correlations over .50 as strong in magnitude (Cohen, 1977, 1988).

The results of the correlation analysis, as displayed in Table 13, illustrate that recall, scored on the original scoring system and the revised scoring system, was highly correlated. When examining the correlation for total recall, the two scoring systems accounted for approximately 74% of the variance in scoring memory recall. This was stable if all cases were included or if cases scoring zero on both scoring systems were excluded.

Table 13

Correlation Between the Revised and Original Scoring Systems

	All cases		Excluding Cases	
Design	*rho*	*N*	*rho*	*n*
1	.77	44	.73	42
2	.81	44	.80	43
3	.70	44	.64	42
4	.83	44	.79	41
Total	.86	44	.86	44

Although there was a very strong correlation between the two scoring systems, difficulties were identified in consistently applying some of the individual items. Therefore, the next step in the development of the revised scoring system involved further refinement of the items.

Comparison Between the Two Scoring Systems After Further Revision

The second stage in the development of the revised scoring system involved reviewing and refining the scoring items. Once a decision was reached on each item in the revised scoring system, 50 clinical protocols for each visual reproduction design (50 immediate recall and 50 delayed recall) were scored on the revised scoring system and the original scoring systems. These clinical protocols were collected from a variety of sources and did not include any of the first sample of 44.

A correlation analysis was conducted between the scores obtained with the original and revised scoring systems to evaluate the similarity of the two systems in measuring design recall (Table 14). Five variables were examined including the score for designs one, two, three and four and the total score. A score was obtained for each of these variables on the revised scoring system and the original scoring system, when rating both immediate and delayed recall. Spearman's correlation coefficients were calculated due to violation of assumptions of normality. When conducting the correlation for delayed recall, a separate analysis was conducted including cases that scored zero on both scoring systems and excluding items that scored zero on both scoring systems.

Table 14

Correlation between the Two Scoring Systems after Further Revision

	Immediate Recall		Delayed Recall			
			All cases		Cases Excluded	
Design	*rho*	*N*	*rho*	*N*	*rho*	*n*
1	.65	50	.82	50	.74	45
2	.69	50	.66	50	.55	46
3	.86	50	.90	50	.90	49
4	.84	50	.83	50	.77	46
Total	.93	50	.94	50	.94	50

When examining the results for immediate recall in Table 14, the correlation between the two scoring systems on each of the designs was very high. The correlation was not as high for designs one and two, but a substantial amount of common variance was present. When examining the total score, the two scoring systems accounted for 87% of the variance.

When examining the results for delayed recall, the correlation between the two sets of scoring systems was also very high. Although coefficients were generally lower when cases scoring zero on both scoring systems were excluded from the analysis, there was not a large difference and there was no difference for the total score (rho= .94). This suggests that including cases that score zero on both scoring systems can inflate the degree of correlation to a modest extent when examining individual designs and a conservative examination of the magnitude of the relationship between the two scoring systems should exclude these cases.

These results supported the notion that the revised scoring system and the original scoring system shared a high degree of association and thereby provided a similar grading of memory functioning. At this point, any differences in the two scoring systems were deemed to be modest. The high association supported the approach to developing the revised scoring system and further analysis of other aspects of the reliability and validity of the revised scoring system was undertaken.

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 revised scoring system and original scoring system 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 each scoring system.

Intra-rater and Inter-rater Reliability

Prior to scoring the revised sample of test protocols, two raters concurrently scored the protocols of a small pilot sample in order to ensure consistency in understanding of the each item. This resulted in some clarification of the terminology and language used in the revised scoring system. However, no changes were made in the meaning of any item and no item was added or deleted at that stage.

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 (the author) rated a new sample of 60 test protocols (the study sample) on both the original scoring system and revised scoring system on two occasions, approximately one to two weeks apart. 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 revised 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 correlation analysis to examine the degree of association between the scores obtained on the two occasions, the percentage of agreement in the score for each design across the two scoring occasions, and the level of agreement on each scoring item. Each of these analyses will be described in more detail now.

Correlation 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 60 protocols scored on two occasions by the one rater. 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. Pearson's correlation coefficients were also calculated for a number of reasons including: 1) to examine if the departures from normality in this sample affected the magnitude of the correlation coefficients, mindful that parametric statistics are considered robust to mild violations of population distribution assumptions, 2) because Pearson's correlations are traditionally used in intra-rater reliability studies, and 3) to compare with intra-class correlations. Intra-class correlations have been used in some reliability studies as they take into account rater agreement rather than rater consistency (McGuire & Batchelor, 1998).

The correlation coefficients for immediate recall are illustrated in Table 15. Both Pearson's and Spearman's correlations were highly similar if not identical. Spearman's correlation coefficients for the revised scoring system were uniformly high, ranging from rho= .94 for Design One to rho= .99 for Design Four. Coefficients for the original scoring system in this sample were generally lower, ranging from rho= .90 for Design Two to rho= .95 for Design Three. However, the differences between two scoring systems were not large and all coefficients were very high. Intra-class coefficients (Appendix F) were very similar to Pearson's coefficients, indicating a high degree of rater consistency and agreement.

Table 15

Correlation Between Scoring of Immediate Recall on Two Occasions (N=60)

	Revised Scoring System		Original Scoring System	
	Pearson's r	Spearman's *rho*	Pearson's r	Spearman's *rho*
Design				
1	.99	.94	.91	.90
2	.97	.98	.88	.90
3	.99	.98	.95	.95
4	.99	.99	.95	.93
Total	.99	.99	.97	.96

When examining delayed recall, two analyses were conducted, one including all of the cases, and another excluding cases that scored zero on both scoring occasions. Spearman's correlations (Table 16), Pearson's correlations (Table 17) and intra-class correlations (Appendix F) were very similar (and almost identical at times). There was minimal or no differences when cases scoring zero on both scoring occasions were excluded from the analysis. When examining the total score for the revised scoring system, there was 98% of common variance across the two scoring occasions. Likewise, there was 94% common variance across the two scoring occasions when examining the total score on the original scoring system.

Table 16

Spearman's Correlation Between the Scoring of Delayed Recall on Two Occasions

	Revised Scoring System				Original Scoring System			
	All cases		Cases excluded		All cases		Cases excluded	
Design	*rho*	N	*rho*	n	*rho*	N	*rho*	n
1	.99	60	.96	34	.98	60	.88	33
2	.99	60	.96	38	.97	60	.86	37
3	.99	60	.99	53	.99	60	.98	49
4	.99	60	.98	44	.97	60	.92	43
Total	.99	60	.99	57	.97	60	.96	54

Table 17

Pearson's Correlations Between The Scoring Of Delayed Recall On Two Occasions

	Revised Scoring System				Original Scoring System			
	All cases		Cases excluded		All cases		Cases excluded	
Design	*r*	*N*	*r*	*n*	*r*	*N*	*r*	*n*
1	.98	60	.97	34	.98	60	.88	33
2	.99	60	.97	38	.97	60	.86	37
3	.97	60	.99	53	.99	60	.98	49
4	.99	60	.99	44	.96	60	.97	43
Total	.99	60	.99	57	.97	60	.99	54

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.

Agreement in total score.

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 implies a high level of agreement). That is, it does not provide direct information as to whether 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, the score agreement across the two scoring occasions for each protocol was collated.

As some error in measurement was to be expected due to random factors, a cutoff level for acceptable agreement in the total score was established. Agreement in the total score within two points was set as an acceptable level of agreement (for e.g. a score of 18 on the first occasion and a score of 20 on the second occasion for the same design would illustrate acceptable agreement). As the total score for each design on the revised scoring system was 20 points, a two-point variation was 10% of the total score on each design. This is double the traditional 5% cutoff often used in psychological research and it was adopted as a reasonable criterion rather than a highly conservative one. This two-point variation was also applied to the original scoring system, thus ensuring a consistent criterion. Therefore, agreement within two points was considered an acceptable level of agreement between the total score obtained on the same design across two separate scoring occasions.

Table 18 shows that when scoring immediate recall on the original scoring system, there was agreement within two points across the two scoring occasions on 100% of the protocols for Design 1, 100% of the protocols for Design 2, 97% of the protocols for Design 3 and 93% of protocols when scoring Design 4. When scoring immediate recall with the revised scoring system, there was agreement within two points on 100% of the protocols for each of the four designs (Table 19). This suggests that there was marginally better total score agreement when the rater used the revised scoring system.

Table 18

Cumulative Percentage of Protocols with Agreement in the Total Score on the *Original Scoring System* for Immediate Recall (*N*=60)

	Level of Total Score Agreement			
Discrepancy	Design 1	Design 2	Design 3	Design 4
Exact Agreement	62%	67%	67%	53%
1 point	100%	92%	95%	80%
2 points		100%	97%	93%
3 points			100%	95%
4 points				97%
5 points				
6 points				98%
7 points				
8 points				100%

Table 19

Cumulative Percentage of Protocols with Agreement in the Total Score on the *Revised Scoring System* for Immediate Recall (N= 60)

	Level of Total Score Agreement			
Discrepancy	Design 1	Design 2	Design 3	Design 4
Exact Agreement	65%	73%	72%	78%
1 point	100%	93%	95%	100%
2 points		100%	100%	

Tables 20 and 21 illustrate total score agreement for delayed recall on the original and revised scoring systems respectively. Table 20 shows that when scoring delayed recall on the original scoring system, there was agreement within two points on 97% of the protocols for Design 1, 100% of the protocols for Design 2, 100% of the protocols for Design 3 and 91% of protocols when scoring Design 4. When scoring delayed recall with the revised scoring system, there was acceptable agreement on 100% of the protocols for Designs 1, 2, 3 and 4 (Table 21). This data supported the results of the correlation analysis and indicated a high level of agreement when scoring the same reproduction across the two scoring occasions.

Table 20

Cumulative Percentage of Protocols with Agreement in the Total Score on the *Original Scoring System* for Delayed Recall

	Level of Total Score Agreement			
	Design 1	Design 2	Design 3	Design 4
Discrepancy	*n*=33	*n*=37	*n*=49	*n*=43
Exact Agreement	67%	62%	86%	53%
1 point	97%	92%	98%	91%
2 points		100%	100%	
3 points	100%			100%

Table 21

Cumulative Percentage of Protocols with Agreement in Total Score on the Revised Scoring System for Delayed Recall

	Level of Total Score Agreement			
	Design 1	Design 2	Design 3	Design 4
Discrepancy	n=35	n=38	n=53	n=44
Exact Agreement	60%	58%	72%	72%
1 point	94%	87%	98%	89%
2 points	100%	100%	100%	100%

It is important to note that a two-point variation in agreement represents more variability in the original scoring system. A two point variation potentially has greater significance when examining the original scoring system as it represents 29% of the total score on Designs 1 and 2. Thus, although the differences between the revised and original scoring systems appear very small when using the two point criterion, the results may reflect a larger difference given that there is a larger range of items on the revised scoring system.

Despite the larger number of items on the revised scoring system, there is a higher correlation between scoring occasions and a higher level of exact score agreement. This suggests that the revised scoring system is a highly reliable system when examining intra-rater reliability.

Level of agreement on individual items.

Although the previous analyses provide information on the consistency of scoring across separate scoring occasions in terms of the total score, they do not provide information about the consistency in which each of the individual items is score. In order to ensure that the other analyses have not obscured problems with scoring individual items, the number of protocols in which the rater scored the same item identically across the two scoring occasions was collated (Appendix G). An analysis of this kind reflects the reliability of scoring individual items.

Because some error in measurement was highly likely, the acceptable level of scoring agreement on each individual scoring item (over the two occasions) was set at 90% of the number of times each item was scored. This is the same criterion, 10% variation, used in the previous comparisons. As sixty protocols were scored, agreement in the score on an individual item on at least 54 of these protocols was considered an adequate level of agreement.

Table 22 illustrates the percentage of scoring items with the same score (0 or 1) across the two scoring occasions on 54 of the reproductions for each of the designs. Of the 41 items in the original scoring system, there was acceptable agreement on 80% of these items when scoring immediate recall and on 98% of the items for delayed recall (Table 22). Out of the 80 items scored for the revised scoring system, there was acceptable agreement on 96% of these items for immediate recall and 100% for delayed recall.

Table 22

Percentage of Items with Acceptable Scoring Agreement

	Original Scoring System % of Items		Revised Scoring System % of Items	
Design	Immediate	Delayed	Immediate	Delayed
1	71%	100%	95%	100%
2	71%	86%	95%	100%
3	78%	100%	95%	100%
4	89%	100%	100%	100%
Total	80%	98%	96%	100%

Poor agreement on an individual item was considered to be reflected by anything less than a 2:1 ratio (meaning that the level of agreement was double the level of disagreement). Therefore any item that had agreement on less than 40 protocols was considered as reflecting poor agreement. On no item was there poor agreement. Appendix G shows a breakdown of each item with the number of protocols that had agreement in score across the two scoring occasions.

Scoring trend.

The number of positive changes and the number of negative changes on individual items from the first to second scoring occasion was examined in order to investigate any significant scoring trend. Table 23 shows the number of positive and negative changes on the original scoring system for immediate and delayed recall. A total of 420 items was scored for Designs 1 and 2 (60 reproductions by 7 items), 540 items were scored for Design 3 (60 reproductions by 9 items) and 1080 items were scored for Design 4 (60 reproductions by 18 items). When scoring immediate and delayed recall on the original scoring system, there did appear to be a slight trend for the rater to become more conservative on the second occasion (Table 23). However, this change is very small.

Table 23

The Number of Positive and Negative Changes from the First to the Second Scoring Occasion using the Original Scoring System

	Immediate Recall		Delayed Recall	
Design	Positive	Negative	Positive	Negative
1	8/420	18/420	7/420	12/420
2	9/420	15/420	11/420	10/420
3	13/540	13/540	4/540	7/540
4	25/1080	34/1080	15/1080	17/1080
Total	55/2460	80/2460	37/2460	46/2460

Table 24 shows the number of positive and negative changes on the revised scoring system for immediate and delayed recall. A total of 1200 items (60 reproductions by 20 items) was scored for each design on the revised scoring system for both immediate and delayed recall.

Table 24

The Number of Positive and Negative Changes from the First to the Second Scoring Occasion using the Revised Scoring System

	Immediate Recall		Delayed Recall	
Design	Positive/1200	Negative/1200	Positive/1200	Negative/1200
1	10	16	11	5
2	18	20	15	14
3	15	13	8	9
4	12	9	7	12
Total	55/4800	58/4800	41/4800	40/4800

The results in Table 24 illustrate that scoring of visual reproduction designs on two occasions by the same rater was highly stable when using the revised scoring system. There is no obvious trend in becoming more conservative or more liberal when scoring with the revised scoring system.

A total of 1200 items where scored for each design on immediate and delayed recall (60 reproductions by 20 items) when using the revised scoring system. The most changes by the rater from the first occasion to the second occasion occurred for immediate recall on Design Two and this was 38 (18 positive, 20 negative). However, this consists of only 3% of the number of items scored indicating a high degree of stability in scoring, illustrating very good stability in scoring.

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 thirty of the test protocols drawn from the sample of 60 noted above. One of these raters was the author and the other the supervisor of this project. Both of these raters were involved in the development of the revised 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 revised scoring system. The variables included in the analysis were the score for each of the four designs and the total score. Examination of inter-rater consistency mirrored that of intra-rater reliability with analyses including correlation analysis, the level of agreement in total score, and the level of agreement on each item.

Correlation analysis.

Spearman's correlations were calculated as the primary measure of inter-rater reliability, however, as outlined for intra-rater reliability Pearson's and intra-class correlation coefficients were also calculated.

The correlation between the scores for the immediate recall is illustrated in Table 25. Generally, both Pearson's and Spearman's correlations were highly similar if not identical. Intra-class coefficients (Appendix H) were identical to Pearson's coefficients. Spearman's correlation coefficients for the four designs were uniformly very high, ranging from rho= .87 to rho= .97 for the revised scoring system and rho= .85 to rho= .94 for the original scoring system when immediate recall was scored. The correlation obtained indicated that there was a very high level of consistency between the two raters on both of the scoring systems, with marginally higher consistency evident on the revised scoring system. As is commonly found the correlation coefficients for inter-rater reliability were generally lower than for intra-rater reliability (Table 15), although the difference in this comparison was very small.

Table 25

Correlation Between Scorers for Immediate Recall (*N*=30)

	Revised Scoring System		Original Scoring System	
Design	Pearson's *r*	Spearman's *rho*	Pearson's *r*	Spearman's *rho*
1	.97	.96	.81	.85
2	.91	.87	.88	.88
3	.96	.94	.83	.85
4	.98	.97	.94	.94
Total	.98	.95	.95	.96

When examining delayed recall, an analysis was conducted including all cases and another analysis was undertaken with reproductions that scored zero on both scoring systems being excluded. In looking at delayed recall, Spearman's (Table 26) Pearson's (Table 27) and intra-class (Appendix H) correlations were highly similar. Spearman's correlation coefficients for the four designs were uniformly very high, ranging from rho= .88 to rho= .94 for the revised scoring system and rho= .70 to rho= .89 for the original scoring system when cases scoring 0 were excluded. When examining the total score for the revised scoring system, there was 96% of common variance across the two scorers. Likewise, there was 94% common variance across the two scorers for the original scoring system.

As shown in Tables 26 and 27, the correlation between raters was uniformly very high for both the original scoring system and revised scoring system. The difference between coefficients when all cases were included or when cases scoring zero by both raters were excluded was also negligible. These results suggested a high degree of consistency between two raters when scoring delayed recall.

Taken together, these results indicate that both the revised and original scoring systems can be quite reliable when used to score Visual Reproduction protocols.

Moreover, in this study the two raters appeared to have applied the same degree of rigor when using both scoring systems.

Table 26

Spearman's Correlation Between Scorers for Delayed Recall

	Revised Scoring System			Original Scoring System		
	All Cases N=30	Cases Excluded		All Cases N=30	Cases Excluded	
Design	*rho*	*rho*	*n*	*rho*	*rho*	*n*
1	.96	.93	25	.87	.70	23
2	.94	.88	24	.93	.84	23
3	.95	.94	27	.93	.89	24
4	.97	.93	24	.94	.89	24
Total	.99	.98	29	.97	.97	29

Table 27

Pearson's Correlation Between Scorers for Delayed Recall

	Revised Scoring System			Original Scoring System		
	All Cases N=30	Cases Excluded		All Cases N=30	Cases Excluded	
Design	*rho*	*rho*	*n*	*rho*	*rho*	*n*
1	.98	.96	25	.91	.70	23
2	.99	.96	24	.96	.89	23
3	.95	.93	27	.94	.91	24
4	.98	.93	24	.95	.92	24
Total	.99	.99	29	.98	.98	29

Agreement in total score.

As was the case with intra-rater reliability, the level of agreement between the raters in the total score obtained for each reproduction was evaluated. The benchmark for acceptable agreement in total score was set at agreement within two points, consistent with the previous comparisons. Tables 28 and 29 illustrate the level of agreement in total score between the two raters for immediate recall on the original and revised scoring systems respectively. Table 28 shows that when scoring immediate recall on the original scoring system, there was agreement within two points on 90% of the protocols for Design 1, 100% of the protocols for Design 2, 90% of the protocols for Design 3 and 86% of protocols when scoring Design 4. When scoring immediate recall with the revised scoring system, there was agreement within two points on 96% of the protocols for Design 1, 93% of protocols for Design 2 and 97% of protocols for Design 3 and Design 4 (Table 29). This represents a marginally higher level of consistency between the two raters on the revised scoring system as compared to the original scoring system. As compared to the results for intra-rater reliability in Tables 18 and 19, there would appear to be marginally lower agreement between two raters as compared to one rater, although the level of agreement is still very high.

Table 28

Cumulative Percentage of Protocols with Agreement in Total Score on the Original Scoring System for Immediate Recall (N= 30)

	Level of Total Score Agreement			
Discrepancy	Design 1	Design 2	Design 3	Design 4
Exact Agreement	37%	50%	37%	33%
1 point	80%	100%	77%	56%
2 points	90%		90%	86%
3 points	100%		93%	96%
4 points			96%	
5 points				100%
6 points			100%	

Table 29

Cumulative Percentage of Protocols with Agreement in Total Score on the Revised Scoring System for Immediate Recall (N= 30)

	Level of Total Score Agreement			
Discrepancy	Design 1	Design 2	Design 3	Design 4
Exact Agreement	53%	30%	20%	33%
1 point	86%	83%	70%	87%
2 points	96%	93%	97%	97%
3 points	100%	96.5%	100%	100%
4 points		100%		

Tables 30 and 31 illustrate the level of agreement in total score between the two raters for delayed recall using the original and revised scoring systems respectively. Table 30 illustrates that there was acceptable agreement on 91% of the protocols for Design 1, 100% of the protocols for Design 2, 96% of the protocols for Design 3 and 76% of the protocols for Design 4 (Table 30). On the revised scoring system, there was acceptable agreement between the two raters, on 92% of the protocols for Design 1, 96% of the protocols for Design 2, 93% of the protocols for Design 3 and 92% of the protocols for Design 4 (Table 31). When comparing inter-rater reliability with intra-rater reliability (Table 20 and Table 21) on the revised scoring system there appeared to be marginally lower agreement on inter-rater reliability.

Table 30

Cumulative Percentage of Protocols with Agreement in Total Score on the Original Scoring System for Delayed Recall

	Level of Total Score Agreement			
	Design 1	Design 2	Design 3	Design 4
Discrepancy	*n*=23	*n*=23	*n*=24	*n*=24
Exact Agreement	57%	43%	38%	40%
1 point	87%	95%	88%	67%
2 points	91%	100%	96%	76%
3 points	100%		100%	84%
4 points				92%
5 points				100%

Table 31

Cumulative Percentage of Protocols with Agreement in Total Score on the Revised Scoring System for Delayed Recall

	Level of Total Score Agreement			
	Design 1	Design 2	Design 3	Design 4
Discrepancy	*n*=25	*n*=24	*n*=27	*n*=24
Exact Agreement	36%	29%	44%	25%
1 point	68%	79%	93%	67%
2 points	92%	96%		92%
3 points	96%	100%		100%
4 points	100%		97%	
5 points				
6 points			100%	

The maximum difference between the scores for each rater on each design can give some indication of the variability in scoring. In examining immediate recall on the original scoring system, Table 28 shows that the maximum difference between the two raters on each of the Visual Reproduction Stimulus Cards were: Design One=3, Design Two=1, Design Three = 6, Design Four =5. Table 29 shows that the maximum difference for the revised scoring system was Design One =3, Design Two =4, Design Three = 3, Design Four =3.

The maximum differences on delayed recall when scored on the original scoring system (Table 30) was Design One =3, Design Two =2, Design Three =3 and Design Four =5. The maximum differences on delayed recall when scored on the revised scoring system was Design One =4, Design Two =3, Design Three =6 and Design Four =3 (Table 31).

The mean score of each rater for each design is shown in Table 32. When examining the original scoring system, the difference between mean scores for total recall between the two raters was 1.1 for immediate recall and 1.3 for delayed recall. When looking at total recall on the revised scoring system, the difference between the mean scores was 0.5 for immediate recall and 0.07 for delayed recall. A comparison between the mean scores by the two raters further supported the notion that the raters were obtaining similar scores for the designs on both scoring systems.

Table 32

Mean Score on each of the Visual Reproduction Designs for each Scorer

	Design 1		Design 2		Design 3		Design 4		Total Score	
Scoring System	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Original										
Immediate										
Scorer 1	4.60	1.61	4.40	1.40	4.30	2.56	8.90	5.52	22.2	8.56
Scorer 2	3.80	1.69	4.50	1.50	4.10	2.23	8.70	5.40	21.1	8.55
Delayed										
Scorer 1	3.57	2.30	3.07	2.12	3.27	2.66	7.33	5.65	17.2	10.8
Scorer 2	3.27	2.23	3.40	2.24	2.83	2.35	6.40	5.32	15.9	10.3
Revised										
Immediate										
Scorer 1	14.7	4.20	13.3	3.12	10.7	3.87	12.0	5.25	50.6	13.2
Scorer 2	14.3	4.07	13.4	3.18	11.5	4.07	11.9	5.44	51.1	13.3
Delayed										
Scorer 1	12.1	6.93	9.77	6.50	8.47	5.13	9.87	6.10	40.2	20.6
Scorer 2	11.6	6.93	9.87	6.50	8.87	5.26	9.80	6.16	40.1	20.9

Level of agreement on each individual item.

The level of agreement between the two raters on each of the 20 individual items scored for each of the four designs was examined in order to evaluate the reliability of scoring each individual item. Thus, the number of protocols on which the two raters scored the same item identically was collated. The acceptable level of agreement over the two occasions was set at 90% of the total number of scoring occasions, in order to make due allowance for error in measurement. This is the same metric, 10% variation, used in the previous comparisons. As thirty protocols were scored by each rater for each of the designs, agreement on 27 of these protocols was considered an adequate level of agreement on each individual item.

Table 33 shows that out of the 41 items scored on the original scoring system, there was acceptable agreement between the two raters on 51% of these items when scoring immediate recall and 71% of these items when scoring delayed recall. As would be expected, the reliability in scoring individual items was marginally better when intra-rater reliability (Table 22) rather than inter-rater reliability was examined.

Table 33

Percentage of Items with Acceptable Scoring Agreement

	Original Scoring System % of Items		Revised Scoring System % of Items	
Design	Immediate	Delayed	Immediate	Delayed
1	57%	71%	90%	95%
2	71%	86%	60%	75%
3	33%	67%	80%	80%
4	50%	67%	75%	80%
Total	51%	71%	76%	835

It is also illustrated in Table 33 that out of the 80 items scored for the revised scoring system, there was acceptable agreement on 76% of these items when scoring immediate recall and 83% of these items when scoring delayed recall. This would appear to reflect a modest trend for the two raters to be more consistent in scoring the individuals items in the revised scoring system.

Poor agreement on individual items was again considered to be reflected by 2:1 agreement/disagreement ratio, therefore any item that had agreement on less than 20 protocols was considered poor agreement. Across the four designs there was poor agreement on only one item of the revised scoring system. This was for Design Three, Item 13. This result suggests that this item could be refined to improve item reliability. Appendix I shows the number of protocols that had agreement on each individual item across the two scoring systems.

Scoring trend.

As for intra-rater reliability, the number of positive changes (from 0 to 1) and the number of negative changes (from 1 to 0) from the first rater to the second rater was collected on each item for each design. Table 34 shows the number of positive and negative changes on the original scoring system for immediate and delayed recall. A total of 210 items was scored for Designs 1 and 2 (30 reproductions by 7 items), 270 items were score for Design 3 (30 reproductions by 9 items) and 540 items for design 4 (30 reproductions by 18 items). When the designs were scored using the original scoring system, there appeared to be an overall trend for more negative changes from the first to the second rater. This appeared to particularly be the case on Designs One and Three. This would suggest that the second rater took a more conservative approach.

Table 34

Number of Positive and Negative Changes on the Original Scoring System

	Immediate Recall		Delayed Recall	
Design	Positive	Negative	Positive	Negative
1	3/210	27/210	4/210	13/210
2	12/210	9/210	15/210	4/210
3	17/270	24/270	5/270	18/270
4	32/540	36/540	14/540	15/540
Total	64/1230	96/1230	38/1230	50/1230

Table 35 shows the number of positive and negative changes on the revised scoring system for immediate and delayed recall. On the revised scoring system, there appeared to be an overall slight trend for the second rater to make more positive changes when scoring immediate recall, particularly on Design Three. This contrasted with more negative changes on Design One.

It must be emphasised that these trends were actually slight since a total of 600 items were scored for each design by each rater (30 protocols scored on 20 items) when rating with the revised scoring system. Therefore a total of 46 changes on Design Three (34 positive, 12 negative) for immediate recall actually represented a very small variation of 7.5%. Hence, even with item agreement, there was a high level of agreement in scoring.

Table 35

Number of Positive and Negative Changes on the Revised Scoring System

	Immediate Recall		Delayed Recall	
Design	Positive/600	Negative/600	Positive/600	Negative/600
1	7	18	13	25
2	25	20	22	19
3	34	12	28	16
4	18	19	17	19
Total	84/2400	69/2400	80/2400	79/2400

Internal Consistency- The Original and Revised Scoring Systems

Measuring the internal consistency of a set of scoring items is also a way of estimating the potential reliability of a scoring system. The internal consistency of the original and revised scoring system was examined by computing Cronbach's alpha's. The internal consistency of the four designs was computed using the combined data from the control and experimental group (n=60). The reliability results are illustrated in Table 36 for each of the Visual Reproduction designs across the revised and original scoring systems. Scoring of delayed recall resulted in marginally larger reliability coefficients. The revised scoring system had larger Cronbach's alpha's for Design 1 and 2 (immediate and delayed recall) and Design 3 (delayed recall).

Table 36

Cronbach's Alpha's for each design on the Revised and Original Scoring Systems

	Immediate Recall	
Design	Revised Scoring System	Original Scoring System
1	.87	.50
2	.73	.65
3	.82	.83
4	.90	.92
Design	Delayed	
1	.98	.90
2	.97	.89
3	.91	.87
4	.94	.94

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 revised scoring system for the Visual Reproduction subtest.

Criterion related validity is based on a test's correlation with other tasks that measure similar and different processes. In this study, concurrent validity, a form of criterion-related validity, was examined. Concurrent validity refers to the correlation of a test with other measures when the other measures are administered at the same time. In this case, the relationship between the revised scoring system and the original scoring system of the Visual Reproduction subtest and another non-verbal memory test (Faces subtest) was examined. Also, the relationship between Visual Reproduction and verbal memory tests (Logical Memory, Verbal Paired Associates) was examined. Concurrent validity was examined in both an experimental and a control group.

This study also examined the construct validity of the revised 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 revised and original scoring systems to discriminate between a group of persons with known cerebral lesions (the experimental 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 an experimental and a control group, the demographic characteristics of these groups were compared.

Demographic Characteristics of the Study Sample

Two groups of thirty participants, one group with documented cerebral lesions and a group with no evidence of cerebral lesions, were compared to see if they were suitably matched groups. Demographic characteristics of individual participants are shown in Appendix J. The results of t-test comparisons between the dependent variable, group membership (control or experimental; male or female) and the independent variables (age, years of education, or gender composition) are shown in Table 37.

Table 37

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

Group	No. S	Variable	*t*	df	*p*
Experimental x Control	*N*=60	Age	.695	58	.490
		Years of Education	-.175	58	.862
Males x Females	*N*=60	Age	-1.925	58	.059
		Years of Education	.538	58	.593
Males	*n*=28	Age	1.470	26	.154
Experimental x Controls		Years of Education	-.686	26	.499
Females	*n*=32	Age	-.253	30	.802
Experimental x Controls		Years of Education	.334	30	.740

Table 37 illustrates that there were no significant differences between the control and experimental groups in terms of age or years of education. A Chi-square analysis also indicated that there was no significant difference in the gender composition of these two groups ($x^2 = .268$, $P>.05$). Table 37 also shows that there were also no significant differences in age or years of education between males and females, between males in the experimental and control groups, or between females in the experimental 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 older. This appeared to be due to two younger men and two older women representing relative outliers.

Employment levels across groups were compared using Mann-Whitney U-Test statistics. There were no significant differences between the employment levels of the participants in the experimental and control groups ($z = -.932$, $P>.05$), between males and females ($z = -1.827$, $p>.05$), between females in the experimental and control groups ($z = .895$, $p>.05$), or between males in the experimental and control groups ($z = .147$, $p>.05$).

Concurrent Validity- Visual Reproduction Revised and Original Scoring System

Relationship Between Memory Measures

Concurrent validity was examined by looking at the correlation between the Visual Reproduction subtest and a number of other external verbal and non-verbal criterion measures in both the control and experimental groups. Scores for the Visual Reproduction subtest (memory for designs, VR) on both the revised scoring system and the original scoring system were correlated with a number of criterion measures including:

- a story memory task (Logical Memory);
- recall of associated and "novel" word pairs (Verbal Paired Associates Hard Pairs and Verbal Paired Associates Easy Pairs); and
- memory for faces (Faces);

Raw scores were used in the analyses as age corrected scores were not available for scoring visual reproductions with the revised scoring system. As was outlined earlier, examination of the data revealed some violation of the assumptions of normality in some of the measures and therefore Spearman's rho correlation coefficients were calculated.

The variables included in the analyses were:

- recall of associated word pairs (Verbal Paired Associates Easy pairs- VPA Easy),
- recall of "novel" word pairs (Verbal Paired Associates Hard pairs- VPA Hard),
- story memory (WMS-R Logical Memory subtest- LM),
- memory for faces (WMS-111 Faces subtest),
- memory for designs (the total score for the revised scoring of the Visual Reproduction subtest- VR Revised- and the total score for the original scoring of the Visual Reproduction subtest-VR Original).

Complete correlation matrices for the Experimental and Control Group are shown in Appendix K.

Correlation between the original and revised scoring systems.

Prior to examining the correlation between memory for designs (VR) and other criterion measures, the relationship between the two scoring systems was examined. A strong correlation between the two systems would suggest that the correlation between memory for designs and other memory measures would be similar regardless of whether the revised scoring system or the original scoring system was used.

In the control group, the correlation between the revised and original scoring systems was rho=.88 for immediate recall and rho=.93 for delayed recall. In the experimental group, the correlation between the revised and original scoring systems was rho=.92 for immediate recall and rho=.93 for delayed recall. Thus, the two scoring systems shared a high degree of common variance.

Correlation between non-verbal memory measures.

When conducting the correlation analyses, it was expected that memory for designs (VR Revised and VR Original) would share a moderate to high correlation with memory for faces (WMS-111 Faces) since both have been used as measures of non-verbal memory. Table 38 illustrates the correlation between the memory for designs and memory for faces in the control group and experimental group, for both immediate and delayed recall. This table shows that there was a moderate correlation between immediate memory for designs and immediate memory for faces in the control group (VR Revised rho=.39 and VR Original rho=.39) and a weak to moderate correlation in the experimental group (VR Revised . rho=35 and VR Original rho=.26).

Table 38

Correlation Between the Recall of Visual Reproductions and Face Memory

	Control Group		Experimental Group	
	Immediate Recall	Delayed Recall	Immediate Recall	Delayed Recall
Variables				
VR Revised & Faces	.39	.58	.35	.44
VR Original & Faces	.39	.44	.26	.48

When examining delayed recall, Table 38 illustrates that there was a moderate to high correlation between memory for faces and memory for designs in the control group (VR Revised rho=.58 and VR Original rho=.44). There was also a moderate relationship between these variables in the experimental group (VR Revised rho=.44 and VR Original rho=.48). The magnitude of the correlation between memory for faces and memory for designs was similar across both scoring systems.

Thus, when examining both immediate and delayed recall for the control group and experimental group, the results are in the expected direction. There is a moderate relationship between the memory of designs and memory of faces. These results suggest that these two tasks measure at least some common aspect of memory functioning, and the nature of the tasks suggests that the common element may well be non-verbal memory.

Relationship between memory for designs and story memory.

It was expected that memory for designs would share only a weak relationship with story memory (LM). The correlation coefficients that illustrate this relationship in the experimental and control group are shown in Table 39.

When examining immediate recall, there was a moderate correlation between the story memory (LM) and memory for designs in the control group (rho=.37 and rho=.38) and a weak correlation in the experimental group (rho=.27 and rho=.23). The magnitude of the correlation between story memory and memory for designs was similar across both scoring systems.

Table 39

Correlation Between the Recall of Visual Reproductions and Story Memory

	Control Group		Experimental Group	
Variables	Immediate Recall	Delayed Recall	Immediate Recall	Delayed Recall
VR Revised & Logical Memory	.37	.24	.27	.58
VR Original & Logical Memory	.38	.21	.23	.55

When examining delayed recall, there was a weak relationship between memory for designs and story memory in the control group (VR Revised rho=.24, VR Original rho=.21). However, in the experimental group, the correlation between story memory and memory for designs was much larger (VR Revised rho=.58, VR Original rho=.55). These coefficients were larger than when immediate memory for designs and immediate story memory were analysed. This result suggests that in the experimental group there may be a stronger correlation between measures of memory, verbal or non-verbal than in the control group, particularly when examining delayed recall.

In the control group, the relationship between the memory for designs and story memory (between non-verbal and verbal memory) was smaller than the relationship between memory for designs and facial memory (non-verbal and non-verbal memory), particularly when examining delayed recall. Thus, delayed memory for designs shares 34% of variance with delayed memory for faces when looking at the revised scoring system (Table 38), but only shares 6% variance with delayed story memory (Table 39).

Conversely, in the experimental group, the relationship between memory for designs and story memory was stronger than the relationship between memory for faces and memory for reproductions, particularly when examining delayed recall (rho

= .58 and .55 versus rho =.44 and .48). Thus, delayed memory for designs shared 19% of its variance with delayed memory for faces when scored according to the revised scoring system (Table 38). In contrast, memory for designs shared 34% of its variance with delayed story memory (Table 39).

These results suggest that there is a moderate to strong relationship between verbal and non-verbal memory measures in the experimental group. However, in the control group, the non-verbal measures share a stronger relationship with each other than with a verbal memory measure, particularly when examining delayed recall.

Relationship between memory for faces and story memory.

In addition to memory for designs, memory for faces was included as a non-verbal memory task. Table 40 illustrates the correlation between memory for faces and story memory. When examining this relationship in the control group, there was generally a weak to moderate relationship between facial memory and story memory. This compares with the previously moderate relationship identified between facial memory and memory for designs in this group.

Table 40

Correlation Between Face Memory and Story Memory

	Control Group		Experimental Group	
Variables	Immediate Recall	Delayed Recall	Immediate Recall	Delayed Recall
Faces & Logical Memory	.29	.34	.42	.44

When examining the relationship between facial memory and story memory in the experimental group, there was a moderate correlation between immediate and delayed memory for faces and story memory (rho=.42 and rho=.44). This is similar to the magnitude of the relationship between memory for designs and memory for faces (Table 38). Again, in the experimental group, there appears to be a moderate relationship between all memory measures, regardless of their content.

Relationship between memory for designs and memory of associated and "novel" word pairs.

Recall of associated and "novel" word pairs was also included in this study as verbal memory measures. The relationship between the recall of associated word pairs with memory for designs is illustrated in Table 41.

When examining immediate recall in the control and experimental groups, there was a moderate correlation between memory for designs and recall of associated word pairs. When examining delayed recall, there was only a weak to moderate correlation between the measures for the control group but a large correlation in the experimental group.

Table 41

Correlation Between the Recall of Visual Reproductions and Recall of Associated Word Pairs

	Control Group		Experimental Group	
Variables	Immediate Recall	Delayed Recall	Immediate Recall	Delayed Recall
VR Revised & Associated Pairs	.38	.16	.34	.51
VR Original & Associated Pairs	.29	.34	.27	.54

The relationship between the recall of "novel" word pairs with memory for designs is illustrated in Table 42. For both immediate and delayed recall, there was a moderate correlation between memory for designs and recall of "novel" word pairs in both groups.

Table 42

Correlation Between the Recall of Visual Reproductions and Recall of Novel Word Pairs

	Control Group		Experimental Group	
Variables	Immediate Recall	Delayed Recall	Immediate Recall	Delayed Recall
VR Revised & Novel Pairs	.56	.43	.47	.42
VR Original & Novel Pairs	.48	.48	.36	.47

The relationship between memory for "novel" word pairs and memory for designs was stronger than the relationship between story memory and memory for designs (Table 39). Thus, both immediate and delayed memory for designs, shared a greater degree of association with the recall of "novel" word pairs, than with story memory

The relationship between memory for designs and memory for faces would seem to be as strong as the relationship between memory for designs and "novel" word pairs. Table 38 showed that there were moderate to large correlations between memory for designs and facial memory (ranging from rho=.35 to rho=.58). Additionally, there were moderate to large correlations between memory for "novel" word pairs in both the experimental and the control groups (ranging from rho=.36 to rho=.56) when examining both immediate and delayed recall. These results suggest that the degree of association between the non-verbal measures (memory for designs and face memory) was similar to the relationship between memory for designs and the recall of "novel" word pairs, that is a non-verbal and a verbal memory measure.

Relationship between memory for faces and memory of associated and "novel" word pairs.

In addition to memory for designs, memory for faces was included as a non-verbal memory task. Table 43 illustrates the correlation between facial memory and memory for associated and "novel" word pairs.

Table 43

Correlation Between Face Memory and the Recall of Word Pairs

	Control Group		Experimental Group	
Variables	Immediate Recall	Delayed Recall	Immediate Recall	Delayed Recall
Faces & Associated Pairs	.34	.01	.05	.25
Faces & "Novel" Pairs	.34	.30	.53	.42

When looking at the relationship between the immediate recall of associated word pairs and memory for faces, there was a moderate correlation in the control group and a weak correlation in the experimental group. These relationships were smaller than those between memory for faces and memory for designs illustrated in Table 38. When examining delayed recalled, the degree of association between faces and associated word pairs decreased in the control group (rho=.01) but increased slightly in the experimental group (rho=.25).

For both immediate and delayed recall, there was a moderate correlation between recall of "novel" word pairs and memory for faces, with this relationship being stronger in the experimental group (Control rho=.34 and rho=.30; Experimental rho=.53 and rho=.42).

In the control group, the magnitude of the relationship between the non-verbal and verbal memory measure (memory for faces and "novel" word pairs), was smaller than that between non-verbal memory measures (memory for faces and memory for designs) when examining delayed recall. In the experimental group, the relationship between memory for faces and "novel" word pairs (Table 43) was similar to the relationship between memory for faces and memory for designs (Table 38).

In the experimental group, the non-verbal measures did not have a higher correlation with each other than with the verbal measures. In the experimental group, all memory measures shared at least a moderate association with each other. Thus, the correlation of different measures of non-verbal memory were no stronger than the correlation between measures of non-verbal memory and verbal memory.

Relationship between verbal memory measures.

The relationship between story memory and the recall of word pairs is illustrated in Table 44. For the control group there was only a small to moderate relationship between recall of associated and "novel" word pairs and story memory. This suggests that there was only a weak relationship in the control group between the measures used to assess verbal memory. Indeed, these coefficients were smaller than the correlation between the verbal measures (story memory and recall of word pairs) and memory for designs. This suggests that in the control group, some of the non-verbal measures have a stronger degree of association with the verbal measures, than the verbal measures have with themselves.

Table 44

Correlation Between Story Memory and Recall of Word Pairs

	Control Group		Experimental Group	
Variables	Immediate Recall	Delayed Recall	Immediate Recall	Delayed Recall
Associated Pairs & Logical Memory	.08	.13	.61	.61
"Novel" Pairs & Logical Memory	.27	.25	.60	.62

When examining the relationship in the experimental group, there was a moderate to large correlation between recall of word pairs and story memory (rho=.43

to rho=.66). These correlation coefficients were larger in the experimental than the control group. This suggests that in the experimental, as compared to the control group, that the measures of verbal memory have a strong degree of association and are measuring at least some similar constructs.

Relationship between memory and cognitive measures.

An examination of the relationship between the Visual Reproduction subtest and the cognitive measures used in this study (Block Design and Vocabulary) was also conducted to provide information about concurrent validity. Table 45 and 46 illustrate the correlation between the memory tasks used in this study and the cognitive measures.

Table 45

Correlation Between Immediate Memory Subtests and the Cognitive Measures in the Control Group

Immediate Recall	V	BD	Delayed Recall	V	BD
Associated Pairs	.23	.31	Associated Pairs	.38	.22
“Novel” Pairs	.47	.53	“Novel” Pairs	.45	.44
Logical Memory	.57	.34	Logical Memory	.38	.22
Faces	.28	.30	Faces	.39	.41
Revised VR	.61	.70	Revised VR	.52	.70
Original VR	.67	.65	Original VR	.54	.71

As can be seen in Table 45, immediate memory for designs had a large correlation with both Block Design (rho=.70 and rho=.65) and Vocabulary (rho=.61 and rho=.67) in the control group. Delayed recall of designs also had a large correlation with Vocabulary, although this was reduced slightly in magnitude as compared with immediate recall and was smaller than the relationship with Block Design. Thus, the relationship between memory for designs (non-verbal memory) and Vocabulary (verbal cognition) was reduced for delayed recall in the control group.

Table 46

Correlation Between Immediate Memory Subtests and the Cognitive Measures in the Experimental group.

Immediate Recall	V	BD	Delayed Recall	V	BD
Associated Pairs	.52	.37	Associated Pairs	.53	.57
"Novel" Pairs	.53	.46	"Novel" Pairs	.60	.45
Logical Memory	.51	.23	Logical Memory	.47	.27
Faces	.15	.40	Faces	.25	.45
Revised VR	.33	.83	Revised VR	.33	.75
Original VR	.30	.80	Original VR	.42	.88

When examining the same relationship in the experimental group (Table 46), it is evident that there was a very high correlation between memory for designs and Block Design (Immediate Recall rho=.83 and rho=.80; Delayed Recall rho=.75 and rho=.88). This was stronger than the relationship seen in the control group. In contrast, the relationship between memory for designs and the Vocabulary subtest was much lower and only moderate in size (Immediate Recall rho=.33 and rho=.30; Delayed Recall rho=.33 and rho=.42). This relationship was smaller than in the control group.

With regard to the other non-verbal memory task, memory for faces, there was a moderate relationship between that subtest and the Block Design subtest in the experimental and control groups. However, this relationship was not much larger than that between memory for faces and the Vocabulary subtest in the control group. It is evident in Tables 45 and 46 that Block Design had a much stronger relationship with memory for designs than with memory for faces.

It would be expected that measures of verbal memory would have a stronger relationship with measures of verbal cognitive ability than non-verbal cognitive ability. Consistent with this, the relationship between the measures of story memory and Vocabulary were generally stronger than the relationship between story memory and Block Design in the control and experimental group.

The relationship between recall of "novel" word pairs and Vocabulary was generally as strong as the relationship between "novel" word pairs and the Block

Design Subtest in both the experimental group and the control groups. Interestingly the relationship between these measures was stronger in the experimental group.

Construct Validity

In addition to looking at the relationship between the Visual Reproduction subtest and other memory measures, validity was also 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 experimental group and the control group.

Experimental and Control Group

Mann-Whitney- U tests were calculated to examine whether there were any significant differences between the control group and experimental group on their memory for designs as scored according to the revised scoring system and original scoring systems. Table 47 illustrates the results of the Mann Whitney U comparison between the experimental and control groups. Mean ranks and sum of ranks are in Appendix L. Due to the large number of comparisons, a conservative significance level was established in order to evaluate the results. A Bonferroni correction was applied and the significance level was set at .005 for each of the 10 comparisons (Revised Scoring and Original Scoring).

When examining the revised scoring system, apart from four comparisons (VR 1 and 2 immediate recall and VR 1 and 4 delayed recall) the scores obtained on all designs were significantly different between the control and experimental groups. For the original scoring system, the scores obtained on the designs were significantly different for five comparisons (VR 1 and 2 immediate recall, VR 2 delayed recall and Total Score on immediate and delayed recall). Of note, when examining the total score, a significant difference between the scores of the control and experimental groups was obtained for both scoring systems. In all cases of significant difference, the control group had a higher mean rank score than the experimental group, indicating that the experimental group performed more poorly than the control group in their memory for designs.

Table 47

Mann Whitney U Comparisons Between the Experimental and Control groups

	z	df	*p*
Revised Scoring			
Immediate			
Design 1	-1.735	58	.083
Design 2	-2.245	58	.025
Design 3	-3.774	58	.000
Design 4	-3.900	58	.000
Total Score	-3.780	58	.000
Delayed			
Design 1	-2.256	58	.024
Design 2	-4.076	58	.000
Design 3	-2.847	58	.004
Design 4	-2.512	58	.012
Total Score	-3.838	58	.000
Original Scoring			
Immediate			
Design 1	-1.145	58	.252
Design 2	-1.408	58	.159
Design 3	-3.648	58	.000
Design 4	-3.698	58	.000
Total Score	-3.924	58	.000
Delayed			
Design 1	-2.327	58	.020
Design 2	-3.740	58	.000
Design 3	-2.541	58	.011
Design 4	-2.380	58	.017
Total Score	-3.346	58	.000

In addition to Mann Whitney U tests, a number of linear Discriminant Function Analyses were carried out to investigate which of the Visual Reproduction scoring systems (original or revised) could best discriminate between the experimental group and control group and to determine the accuracy of participant classification using the Visual Reproduction scoring systems. Although not all variables fulfilled the assumptions underlying this method of statistical analysis, it was considered to be robust to violations of these assumptions, particularly when each group has the same number of subjects (Tabachnick & Fidell, 1996). Thus, Discriminant Function Analyses were carried out, although it is acknowledged that some caution would be needed in interpreting the results.

When the immediate recall and delayed recall scores on the original scoring system for each of the four designs were entered into a Discriminant Function Analysis, the resulting equation was significant (x (8) =34.392, $p>.01$). This equation resulted in the correct classification of 90% of the experimental group and 86.7% of the control group. A significant equation also resulted when only the total scores for immediate and delayed recall were entered as the predictors. This function correctly classified 78.3% of the subjects (73.3% of the experimental group and 83.3% of the control group). Thus, classification was better when all the designs were entered separately, rather than as a total score.

When the immediate recall and delayed recall scores on the revised scoring system for each of the four designs was entered into a linear Discriminant Function Analysis, the resulting function was significant (x (8) =34.118, $p>.001$) with all designs contributing significantly to the equation. The function correctly classified 86.7% of the subjects in the experimental and control groups. When the total scores for immediate and delayed recall on the revised scoring system were entered into a linear Discriminant Function Analysis, the equation was again significant (x=19.392 (2), $p>.01$). This equation correctly classified 66.7% of the subjects in the experimental group and 76.7% of the subjects in the control group. The classification of subjects was superior when scores for each of the designs were entered into the analysis, rather than the total score for the four designs.

Left and Right Group

Mindful of the association between verbal memory and the left hemisphere of the brain and the association between non-verbal memory and the right hemisphere of the brain, construct validity could be further supported by analysing the difference between groups with circumscribed left and right brain lesions on the Visual Reproduction subtest. It would be expected that scores on the Visual Reproduction subtest would be different between two groups with clearly lateralised brain damage and that the group with right hemisphere damage would perform more poorly on the Visual Reproduction subtest.

In the sample collected for this study, only 14 out of the 30 experimental participants had clearly lateralised damage (5 left, 9 right). Hence, this relatively small sample was supplemented with Visual Reproduction protocols from the same rehabilitation population from which the experimental group was derived. Visual Reproduction protocols were selected from archives if an individual had a confirmed circumscribed lesion that was clearly lateralised. Protocols of a further 9 left hemisphere and 7 right hemisphere participants were obtained, resulting in a total sample of 16 participants with right sided lesions and 14 participants with left sided lesions (Details of these participants are shown in Appendix M).

Mann Whitney U statistics were used to compare the two groups due to violations of normality in some of the variables. Results of the Mann Whitney U tests were illustrated in Table 48. Mean ranks and sum of ranks are in Appendix N. Due to the large number of comparisons, a conservative significance level was used to evaluate the comparisons between the two groups. A Bonferroni type correction was applied and the significance level was set at .005 for each of the 10 comparisons performed on the Revised Scoring System and the Original Scoring System.

Table 48

Mann Whitney U Comparisons Between the Left and Right Hemisphere Groups

	z	df	*p*
Revised Scoring			
Immediate			
Design 1	-1.136	28	.275
Design 2	-1.029	28	.313
Design 3	-2.880	28	.003
Design 4	-2.735	28	.005
Total Score	-2.812	28	.004\
Delayed			
Design 1	-2.634	28	.019
Design 2	-1.057	28	.335
Design 3	-3.026	28	.002
Design 4	-3.098	28	.002
Total Score	-4.142	28	.000
Original Scoring			
Immediate			
Design 1	-1.368	28	.193
Design 2	-.106	28	.918
Design 3	-2.004	28	.047
Design 4	-3.093	28	.001
Total Score	-2.897	28	.002
Delayed			
Design 1	-2.423	28	.031
Design 2	-.876	28	.448
Design 3	-2.702	28	.008
Design 4	-3.417	28	.001
Total Score	-4.397	28	.000

The results in Table 48 demonstrate that there was no significant difference between the scores obtained by the two groups for immediate recall on Designs 1 and 2 and delayed recall on Designs 1 and 2. When examining the revised scoring system, there was a significant difference between the left and right groups when the immediate reproductions and the delayed reproductions of Design 3 and 4 (revised scoring) were scored. On the original scoring system, there was also a significant difference between the groups when the immediate and delayed reproductions on Designs 4 were scored.

When looking at total recall, there was a significant difference between the left and right lesioned group for immediate and delayed recall on the revised scoring system and the original scoring system. In all significant cases, the left group performed better than the right group. Thus, the group with right hemisphere deficits appeared to have poorer non-verbal memory overall (as represented by memory for designs) than a group with right hemisphere deficits.

In addition to Mann Whitney U tests, a number of Discriminant Function Analyses were carried out to investigate which of the Visual Reproduction scoring systems (original or revised) could best discriminate between the left and right groups and to determine the accuracy of participant classification with the Visual Reproduction scoring systems. This analysis was explorative rather than definitive given that not all the variables fulfilled the assumptions underlying this method of statistical analysis. Caution regarding results is also necessary given the limitations of this technique when the classification success rate of the function is not tested in a different sample from the one in which the function was generated (Tabachnick & Fidell, 1996).

In the first analysis, immediate recall and delayed recall scores on the original system for each of the four designs was entered into a linear Discriminant Function Analysis. The resulting function was significant ($x(8) = 37.364$, $p > .001$), with all designs contributing significantly to the equation. The function correctly classified 92.9% of those with left hemisphere damage and 93.8% of those with right hemisphere damage. When the total score for the original scoring system were entered into a linear Discriminant Function Analysis the equation was significant ($x(2) = 28.850$, $p > .001$). This equation classified 78.6% % of the left hemisphere group and 93.8% of the right hemisphere group. Thus, the ability to correctly classify participants was enhanced by including each design in the Discriminant Function Analysis, rather than just entering the total scores for immediate recall and delayed recall.

When the immediate recall and delayed recall scores on the revised scoring system for each of the four designs were entered into the Discriminant Function Analysis, the function was again significant (x (8) =30.821, p>.01). This equation correctly classified 100% of the left hemisphere group and 87.5% of the right hemisphere group. When the total score for immediate and delayed recall were entered as the predictors, the function was again significant and correctly classified 92.9% of the left hemisphere group and 81.3% of the right hemisphere group. Thus, classification rates were marginally better when each design was included as a predictor.

Non-Verbal Memory Index

As noted above, memory for designs as scored on the original scoring system or the revised scoring system was significantly poorer in a group with right lateralised lesions as compared to left lateralised lesions. Hence, the Visual Reproduction subtest has the potential to provide a measure of non-verbal memory. However, it is likely that verbal memory processes contribute to performance on this task as well, since there is a correlation between this task and verbal memory performance.

The development of a non-verbal memory index was explored in this study. This Index was based on the identification of particular items on the revised scoring system that appeared to reflect the contribution of non-verbal memory. In order to investigate this, performance on each item was compared in the groups with left and right lateralised damage.

Mann Whitney U comparisons were conducted between the lateralised left and right hemisphere groups on each individual item, both on the revised and the original scoring systems. Table 49 shows the items where a significant difference was found between the two groups using the original scoring system. Table 50 lists the items where a significant difference was found between the groups using the revised scoring system. In all cases where a significant difference was found, the right lateralised group performed more poorly than the left lateralised group. The complete data is provided in Appendix O.

Table 49

Items that were significantly different between the two groups on the original scoring system

	Design 1	Design 2	Design 3	Design 4
Immediate Recall	No items	No items	No Items	Items 3, 4, 7, 9 Items 10, 17, 18
Delayed Recall	No items	No Items	Items 5, 8	Items 1, 3, 4, 9 Items 10, 15, 18

Table 50

Items that were significantly different between the two groups on the revised scoring system

	Design 1	Design 2	Design 3	Design 4
Immediate Recall	No items	No items	Items 3, 7	Items 6, 8
Delayed Recall	Items 1, 4, 5, 7 Items 12, 13, 14, 15, 16, 20	No Items	Items 1, 8, 9 Items 10, 11, 15	Item 2, 3, 4, 6, Item 8, 17, 18

Items on the original scoring system that discriminated between the two groups were found for Design 4 on both immediate and delayed recall and on Design 3 for delayed recall (Table 49). Items on the revised scoring system that discriminated between the two groups were found on Designs 3 and 4 (immediate and delayed recall) and on Design 1 for delayed recall (Table 50).

The items for which there was a significant difference between the left and right lateralised groups could potentially provide an index of non-verbal memory function. Given the association between non-verbal memory and the right hemisphere of the brain, poor performance by a group with right hemisphere compromise on these items could indicate that these items are sensitive to non-verbal memory. However, exclusion of confounding factors was a important step in producing this index. Thus, when deriving this index a number of exclusionary criteria were used. Firstly, only items that were found to discriminate between the groups on delayed recall were used since immediate recall might be confounded by factors other than memory. Secondly, any item that discriminated between the two groups on delayed recall was excluded if it also discriminated on immediate recall as performance on this item could be potentially confounded by other processes such as visuo-constructional difficulties. Thirdly, any item that was performed poorly by both groups was excluded, even if there was a significant difference between groups. For example Item 15 on Design 3 in the revised scoring system was performed poorly by both the left and right lesion groups).

With regard to the original scoring system, five items remained as a potential index of non-verbal memory processes (Items 5 and 8 on Design 3 and Items 1, 15 and 18 on Design 4). In order to evaluate if these five items could be used as a reliable index of non-verbal memory functioning, a linear discriminant function analysis was conducted to determine the accuracy of classification with the two groups who had lateralised cerebral damage based on these five items. Item 8 on Design 3 failed the tolerance test but all other items were entered into the analysis. The resulting equation was significant (x (4)= 15.216, p=.004). This equation correctly classified 85.7% of the participants into the Left group and 81.3% of the participants into the right hemisphere group. Two of the members of the left hemisphere group were incorrectly classified, whilst 3 members of the right hemisphere group were incorrectly classified.

After exclusion of Items 3, 7 and 15 from Design Three and of Items 6 and 8 of Design Four from the revised scoring system, a total of 20 items remained across the designs. These 20 items, as listed in Table 51 were examined as an index of non-verbal memory functioning via a linear discriminant function analysis.

Table 51

A 20 item Scale of Non-Verbal Memory Derived from the Revised Scoring System

DESIGN	ITEMS
1	Items 1, 4, 5, 7, 12, 13, 14, 15, 16, 20
3	Items 1, 8, 9, 10, 11
4	Items 2, 3, 4, 17, 18

A linear discriminant function analysis was conducted to determine whether these 20 items were able to discriminate between the left and right groups and to determine the accuracy of participant classification using the index. Items 7, 13 and 14 from Design One and Item 9 from Design Three failed the tolerance test and were not entered into the function. The remaining 16 items were entered into the analysis with the resulting equation being significant (x (16)=42.331, $p>.01$). This function correctly classified 100% of the left hemisphere group and 100% of the right hemisphere group. Thus, the 16 items that were entered into the discriminant analysis item scale were able to correctly classify all the left and right hemisphere participants. Hence this group of 16 items from the revised scoring system could be used as an Index to identify non-verbal memory dysfunction. The items in the proposed Index are illustrated in Table 52.

Table 52

A Potential Index of Non-Verbal Memory

Design	Item	Description
Design 1	1	There are at least two continuous lines or 4 lines eminating from a central point or figure.
	4	The lines that intersect or emanate from the central point do not form angles < 45.
	5	Lines or radial spokes are similar in length. The shorter line or spoke must be at least 75% of the length of the longest line (regardless of where they intersect).
	12	Exactly 4 discrete figures are present. The figures do not share a border with an external square.
	15	All four figures are correctly position near the end of the appropriate line.
	16	All four figures touch one line at the endpoint of the line.
	20	No extra elements.
Design 3	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. Treat as a large square if there are small gaps in the lines. If there is more than one large outside square, treat the outermost square as the large figure. If there is any doubt, score to maximise scoring.
	8	Two to four smaller figures are present with or without a major figure bordering them.
	10	Smaller figures are distinct shapes.
	11	Each of the smaller figures is divided into four parts, or there are four shapes in each quadrant in a 2 x 2 matrix.
Design 4	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 than the parallel side.
	3	One or more three to six sided figures is adjacent to the large rectangle (sharing a border is acceptable). If there is no large rectangle, there are one or more 3-6 sided figures present.
	4	The smaller figure/s in item three are separate from each other and from the major figure.
	17	The smaller figure is located to the right of the figure in 10, or to the right of some rectangle if no 10.
	18	The smaller figure is located at or near to the centre of the right border of the figure in item 10. The smaller figure must be located within + or - 30 degrees of the arc of the figure in 10 (the figure can be inside another).

Figure 4 shows the distribution of scores on the Index for each participant in the left lesion and right lesion groups. As shown in this figure, no participant in the right lesion group had a score higher than eight. No participant in the left group had a score lower than seven. In contrast, only two of the participants in the right lesion group scored more than seven points on the Index, with one participant scoring eight. Only three of the participants in the left group scored eight or less than on the Index. Discrimination between the group may be hardest between the scores of seven and eight, where participants from both groups obtained scores. Conversely, a score below five on the Index may be indicative of non-verbal memory dysfunction.

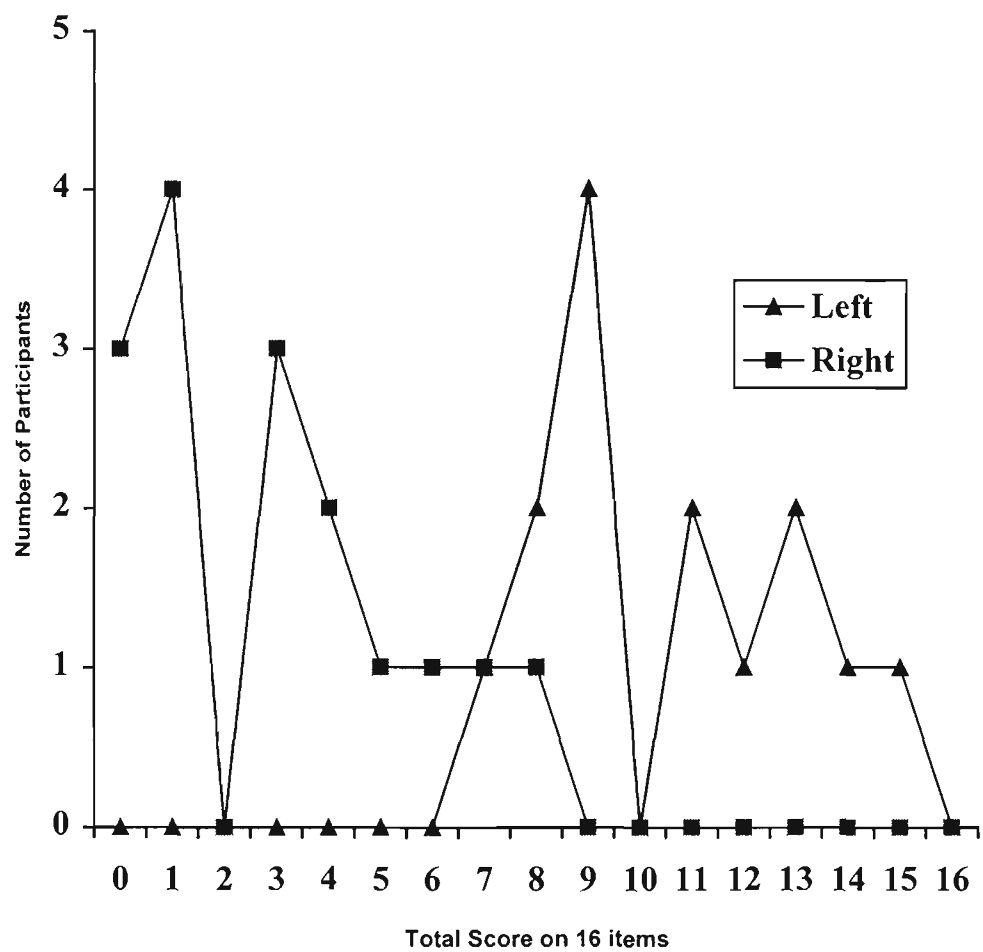


Figure 4: Distribution of scores on the 16 item Index for the left and right groups.

Normative Data for the Visual Reproduction Revised Scoring System

As the sample in this study was small and not stratified, it was not possible to directly derive any normative data for scoring by the revised scoring system. However, examining the median scores on each of the scoring systems in the control group could serve as a preliminary comparison between the two scoring systems.

Before comparing the scoring systems, it was considered important to establish if the control group in this study were different from the standardisation sample of the WMS-R. The general intellectual level of the control group was broadly examined via performance on two WAIS-R subtests that have high loading on general ability. The mean score for the control group on the Vocabulary and Block design subtests are shown in Table 53 and the data indicates that the group was of average intelligence.

Table 53

Mean Scores on Cognitive Measures for the Control Group

	Mean	Standard Deviation
Vocabulary	9.87	1.83
Block Design	10.37	2.58

All scores on the original scoring system were then converted to percentiles according to the norms in the WMS-R manual. The median score of immediate recall was the 69th percentile and the median score on delayed recall was the 54th percentile. Thus, the average intelligence sample had an average level of performance on the Visual Reproduction subtest as scored on the original scoring system.

Table 54 provides the median scores on each of the visual reproduction designs on the revised and the original scoring system. This table illustrates that the median score for both scoring systems is generally positively skewed.

Table 54

Median Scores for each Design in the Control Group

	Immediate Recall		Delayed Recall	
Design	Original Scoring System	Revised Scoring System	Original Scoring System	Revised Scoring System
1	5	17.5	4.5	16.5
2	5	15.0	4	14.0
3	6	14.0	5	13.0
4	12.5	16.0	8	13.5
Total	29.5	61.5	21	50.5

Table 55 illustrates Interquartile ranges for each scoring system in the control group for immediate recall.

Table 55

Interquartile Ranges for each Design for Immediate Recall in the Control Group

	Immediate Recall							
	Original Scoring System				Revised Scoring System			
	Interquartile Range				Interquartile Range			
Design	25	50	75	Range	25	50	75	Range
1	4	5	6	2	15.75	17.5	19	3.25
2	4	5	6	2	12	15.0	17	5
3	4.75	6	8	3.25	11.75	14.0	17.25	6
4	9.75	12.5	16	6.25	14	16.0	17	3
Total	25.5	29.5	34.25	8.75	59	61.5	66.5	7.5

Table 55 shows that 50% of the scores obtained in the control group on each of the designs fell within similar range for the original and revised scoring systems, particularly when examining delayed recall. This data could potentially provide the basis for a score conversion, with performance at the 75th percentile on the revised scoring system corresponding to a score at the 75th percentile on the original scoring system.

Table 56 illustrates the Interquartile ranges for each design for delayed recall. When examining delayed recall, 50% of the scores obtained in the control group fell within a larger range on the revised scoring system as compared to the original scoring system. There was also a wider range of potential scores below the 25th quartile for the revised scoring system on designs two, three and four. When looking at the total score for delayed recall, there was a greater range of potential scores below the 25th quartile for the revised scoring system (35.5 as compared to 13.75). There was also a greater range of potential scores above the 75th quartile for the revised scoring system (17.25 as compared to 19.25).

Table 56

Interquartile Ranges for each Design for Delayed Recall in the Control Group

	Delayed Recall							
	Original Scoring System				Revised Scoring System			
	Interquartile Range				Interquartile Range			
Design	25	50	75	Range	25	50	75	Range
1	0	4.5	6	6	0	16.5	19	19
2	3	4	6	3	11	14.0	16.5	5.5
3	3	5	7.25	4.25	9	13.0	16.25	7.25
4	4.75	8	14.25	9.45	8.5	13.5	17	9
Total	13.75	21	30.25	16.5	35.5	50.5	60.75	25

Due to the small sample size in this study it was not possible to derive any age corrected normative data for the revised scoring system. Thus, in order to provide some preliminary data on converting scores on the revised system to scores on the original system, the raw scores (for total recall) were plotted for immediate and delayed recall. Figure five illustrates the relationship between raw scores on the revised scoring system and the original scoring system for immediate recall and Figure six illustrates the relationship between delayed recall raw scores.

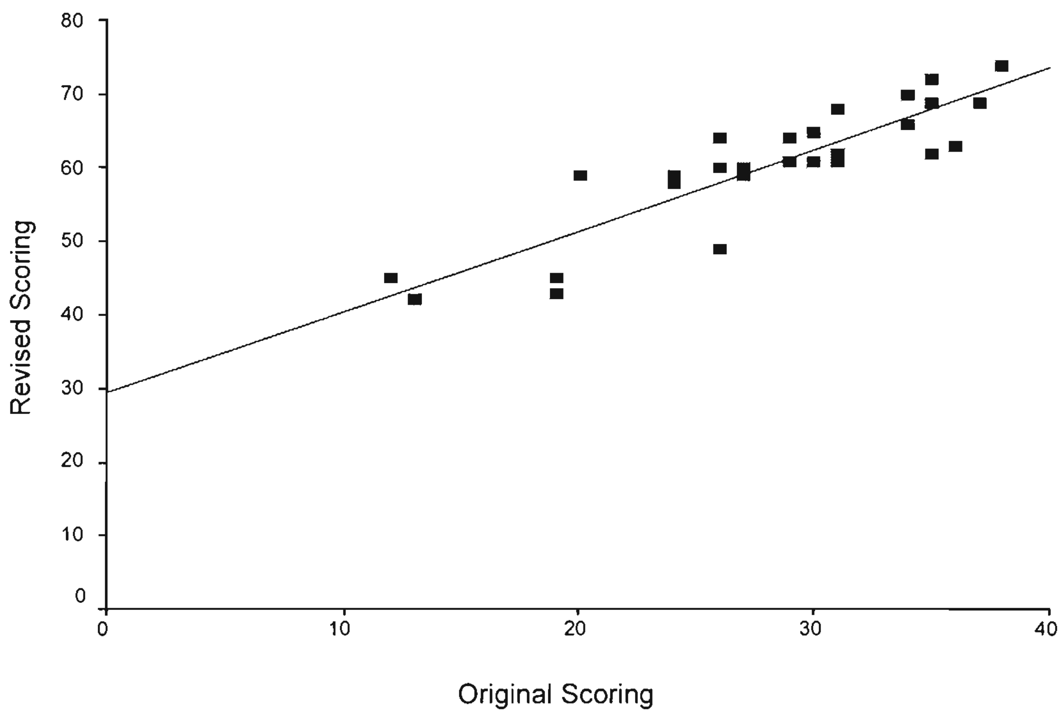


Figure 5: Immediate Recall- Scatterplot of the total raw scores on the revised and original scoring systems

Figure 5 shows that there was a strong linear relationship between the score derived from the original scoring system and the score obtained for the revised scoring system for immediate recall. This is consistent with the very high correlation found between the two scoring systems (rho=.88).

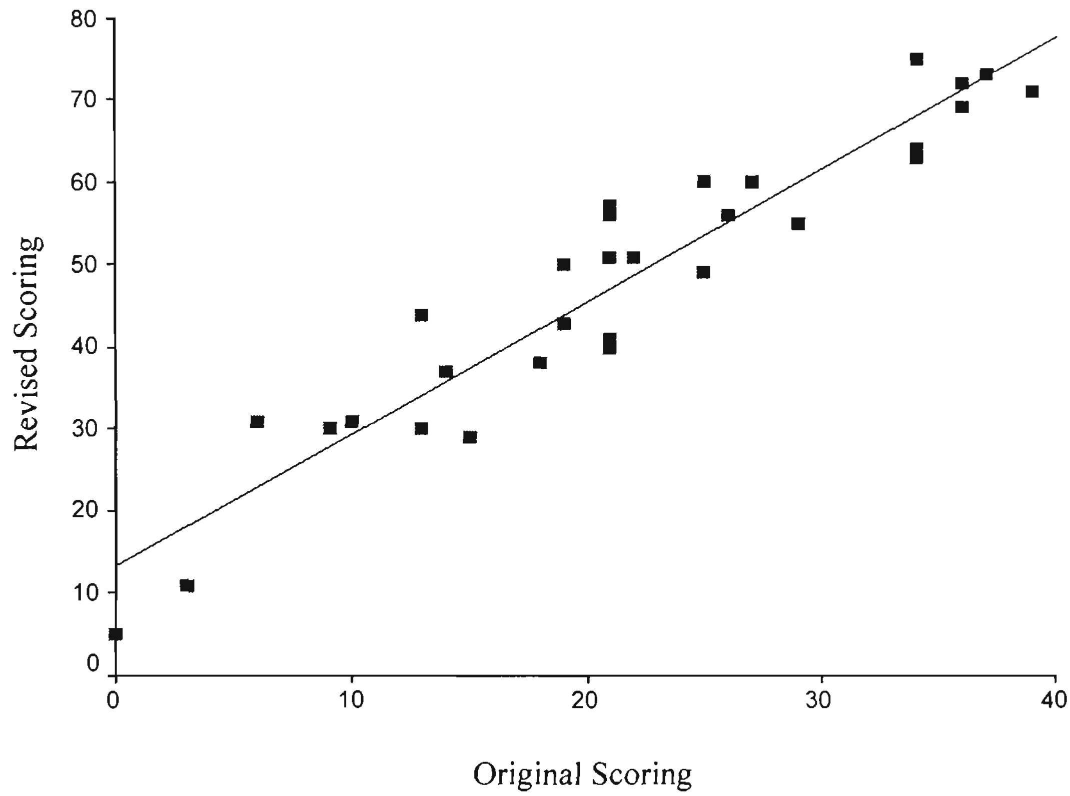


Figure 6: Delayed Recall- Scatterplot of the total raw scores on the revised and original scoring systems.

Figure 6 also shows that there was a strong linear relationship between the delayed recall scores derived from the original scoring system and the revised scoring systems. Again, this is consistent with the very high correlation found between the two scoring systems (rho=.93).

Further to this, the relationship between raw scores (for total recall) on the revised scoring system and percentile scores on the original scoring system was examined for those participants in the control group aged between 70-74. This is shown in Figure 7.

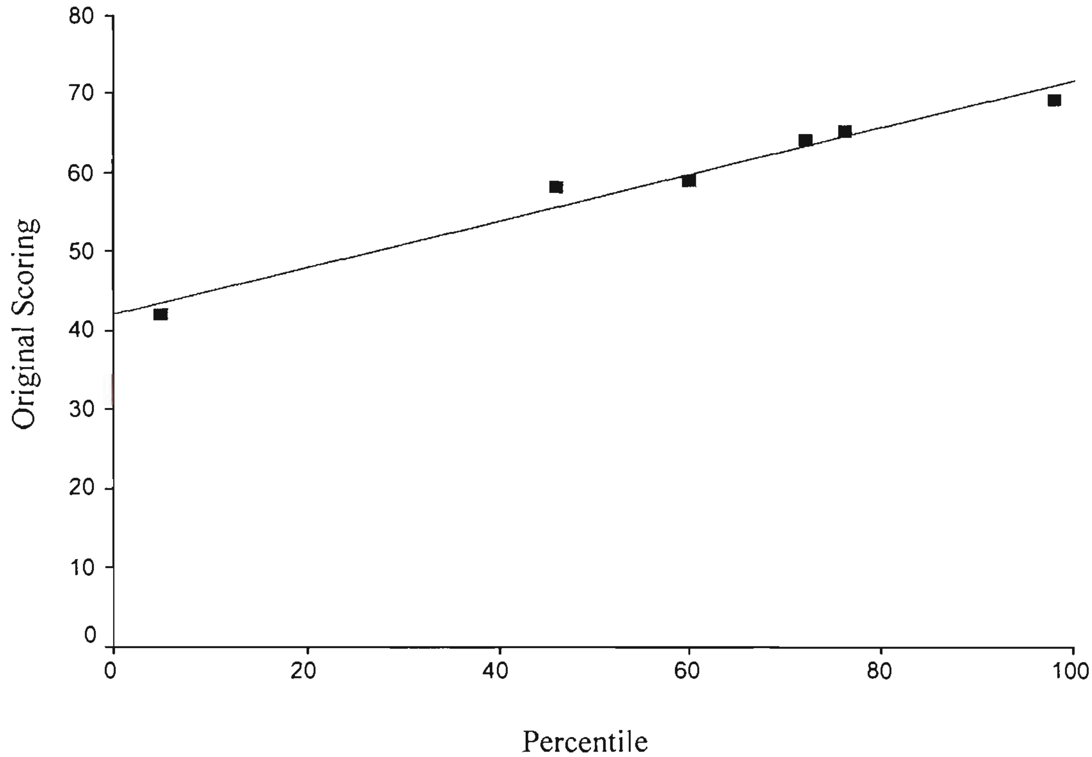


Figure 7: Immediate Recall- Scatterplot of the total raw score on the revised scoring system and the percentile score on the original scoring system.

Figure 7 illustrates the relationship between the raw scores on the revised scoring system and the corresponding percentiles on the original scoring system for immediate recall. This figure also illustrates a strong linear relationship between the original and revised scoring systems.

Figure 8 illustrates the relationship between the raw scores on the revised scoring system and the corresponding percentiles on the original scoring system for delayed recall.

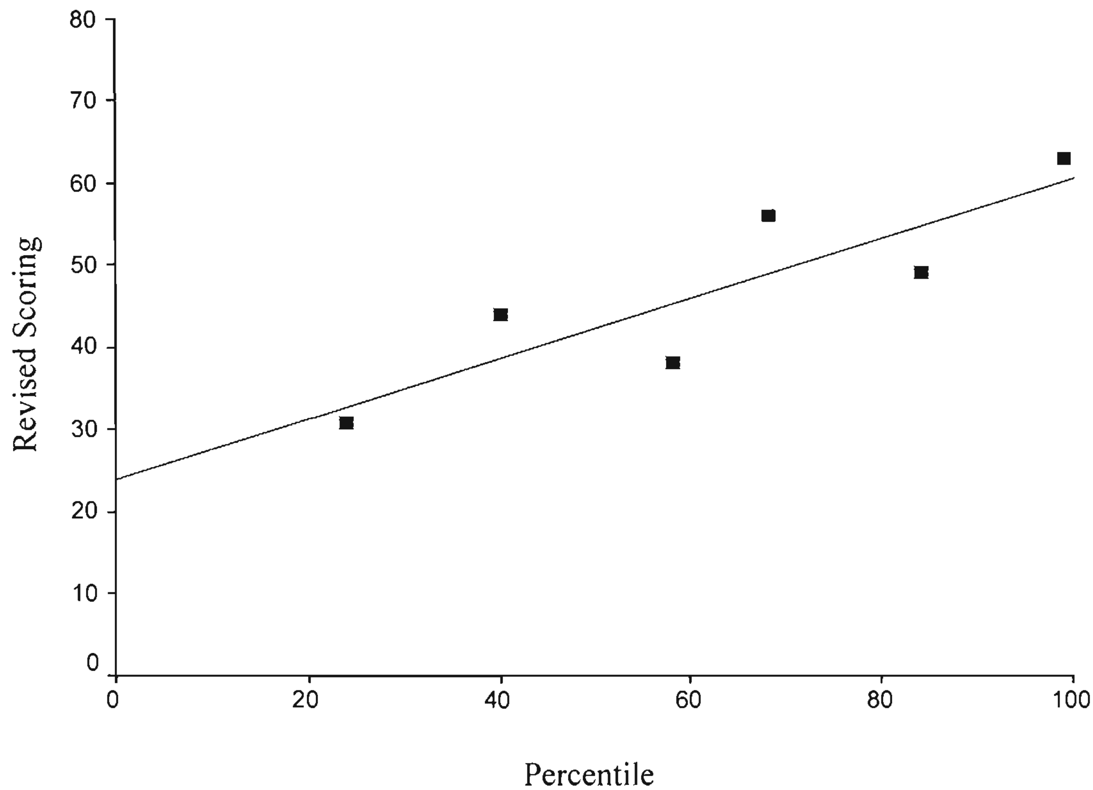


Figure 8: Delayed Recall- Scatterplot of the total score on the revised scoring system and the percentile score on the original scoring system.

Figure 8 figures suggests a clear linear relationship between the scores obtained on the two scoring systems. This strong linear relationship would be expected given the high correlation between the scores obtained on the two scoring systems. This suggests that with a larger sample of participants there is considerable promise for developing an equation predicting the score on the revised scoring system from the old scoring system and vice versa.

Range of Scores Obtained in the Combined Study Sample

Table 57 illustrates the range of scores obtained for immediate recall on the original and revised scoring systems. Floor and ceiling effects are evident when examining the range on the original scoring system. There is also a wider range of potential scores below the 10th percentile on the revised scoring system. Increasing the number of items on the revised scoring system appears to have resulted in a wider range of scores across each of the designs. Whilst the range for the original scoring system generally encompasses the number of items for each design, the range is still narrower than the revised scoring system, because of the reduced number of items on the original scoring system.

Table 57

Percentiles and Range for Immediate Recall in the Combined Study Sample (N=60)

	Revised Scoring System							
	Percentiles					Range		
Design	10	25	50	75	90	Min	Max	Range
1	11	14	17	18	19	5	19	14
2	10	11.25	13	16	18	7	20	13
3	7	9	12	16	18	3	20	17
4	6	8.25	14.5	17	18	0	20	20
Total	41	45	58	64.75	69	20	74	54
	Original Scoring System							
	Percentiles					Range		
Design	10	25	50	75	90	Min	Max	Range
1	3	4	5	6	6	0	7	7
2	2	3.25	5	6	7	1	7	6
3	2	3	5.5	7	8	0	9	9
4	1	6	10.5	14.75	16	0	18	18
Total	13	17	24	31	35	5	38	33

Table 58 illustrates the range of scores obtained for the original and revised scoring systems on delayed recall. Whilst ceiling effects appear to be generally limited, floor effects are apparent on both scoring systems. Again, there is a wider range of scores produced by the revised scoring system by virtue of the fact it contains more items than the original. Increasing the number of items on the revised scoring system appears to have resulted in a wider range of scores on delayed recall across each of the designs, particularly for the first three designs.

Table 58

Percentiles and Range for Delayed Recall in the Combined Study Sample (N=60)

	Revised Scoring System							
	Percentiles					Range		
Design	10	25	50	75	90	Min	Max	Range
1	0	0	13	17.75	19	0	20	20
2	0	0	11	15	18	0	19	19
3	0	7.25	10.5	14.75	17	0	20	20
4	0	0	10	15.75	17.9	0	20	0
Total	5	23.5	37.5	55.75	63.9	0	75	75
	Original Scoring System							
	Percentiles					Range		
Design	10	25	50	75	90	Min	Max	Range
1	0	0	3	5	6	0	7	7
2	0	0	3	5	6	0	7	7
3	0	2	3.5	7	8	0	9	9
4	0	0	6.5	12	15.9	0	18	18
Total	0	9	18	25	34	0	39	39

Clinical Usefulness

Cued Format

In the elaborated administration of the Visual Reproduction subtest used in this study a cue was provided if a participant had no recall of a particular design on delayed recall. The benefit of this cue was then evaluated. A total of 24 participants in the experimental group (80%) were provided with a cue for at least one design. Cues were required on more than one design for 14 participants (46% of the group) and two participants required cueing for all four designs because they had no free recall. Of the 24 experimental participants provided with cues, 66% were able to benefit from cueing on at least one design. Eight participants were unable to benefit from cueing on any design.

With respect to the control group, 14 participants required cueing (47% of the group). Cues were provided on more than one design in three cases (10% of the group). Of these 14 control participants, 86% were able to benefit from cueing on at least one of the designs they received a cue on.

Taken together, the results from the experimental and control group suggested that the provision of cues when a participant had no free delayed recall for a design facilitated recall in the majority of cases.

Tables 59 and 60 illustrate the number of times that cueing was provided for each of the four designs and whether these facilitated recall in the experimental and control groups respectively. The average score gain on the revised scoring system when a cue was successful in eliciting recall is also provided. These tables indicate that when a cue was provided for design one or two, participants recalled a substantial amount of the design. In many cases this additional recall was more than 50% of the points allocated to a design.

Table 59

Benefit from Cueing on the Visual Reproduction Task in the Experimental Group

	Experimental Group		
Design	No of times Cues given	Percentage with some recall	Average Score Gain
1	14	57	11.75 Points
2	18	78	12.63 Points
3	5	0	
4	11	0	

Table 60

Benefit from Cueing on the Visual Reproduction Task in the Control Group.

	Control Group		
Design	No of times Cues provided	Percentage with some Recall	Average Score Gain
1	11	91	16 Points
2	3	33	11 Points
3	0	0	-
4	4	50	6.5 Points

Taken together, these results suggest that for some of the participants in this study using a cued format could facilitate retrieval. However, not all the participants benefited from a cue.

Multiple Choice

Following delayed recall and in some cases cued recall, each participant was given a multiple choice recognition task for each of the four designs. Table 61 illustrates the frequency of scores on the Multiple Choice task. Only two people in the control group failed to have perfect recognition, suggesting that this task was relatively easy for an unimpaired population. Ten subjects in the experimental group had less than perfect recognition.

Table 61

Frequency of Scores on the Multiple Choice task

	Frequency		
Multiple Choice Score	Group	Experimental	Control
0	1	1	0
1	4	3	1
2	3	3	0
3	4	3	1
4	48	20	28

Many of the control participants performed well on the multiple-choice task but this was to be expected when many were able to recall these designs after a delay or with assistance of cueing. To establish the clinical utility in using this procedure it is important to establish whether the multiple-choice task assisted recall in participants who did poorly on delayed/cued recall. Of the 24 participants in the experimental group who were provided with a cue on one or more design, 33% (8) where unable to benefit from cueing on any design they had not freely recalled. The majority of these participants (7 of 8) were able to recognise at least one of the designs from a recognition format.

It was potentially informative to explore whether participants who had no recall, either on delay or after a cue, were able to subsequently recall the design when given a multiple choice recognition task. Thus, the number of participants who scored zero on delayed/cued recall for each of the designs was collected. This was then compared with each participant's multiple-choice recognition of the design/designs of which they had no recall.

The relationship between scoring zero on delayed and cued recall and subsequent multiple choice recognition for each of the designs is shown in Table 62. Overall, when a participant failed to recall a design, there was a trend for them to be able to recognise the design when given a recognition format particularly on designs 1 and 2. In terms of clinical utility this suggests that the multiple choice task is potentially of considerable use in evaluating aspects of preserved memory function not shown by free recall performance.

Table 62

Multiple Choice Recognition of each Design in Participants with no delayed or cued Recall

Design	Percentage of total sample with some free recall	Percentage of participants with no delayed recall who benefited from a cue	Percentage of participants with no free or cued recall who benefited from multiple choice
1	56%	72%	63%
2	63%	57%	80%
3	88%	0%	20%
4	56%	13%	50%

When analysing this data in terms of participants rather than designs, 13 experimental participants had no recall of one or more of the designs on delayed or cued recall. Of these 13, 11 (76%) were able to recognise one or more of the designs that they had no prior delayed or cued recall of. Of these 11, 72% were able to correctly identify all the designs for which they had no delayed recall or cued recall. These results suggest that a significant percentage of those who had no delayed recall and no cued recall were able to recognise the design when given a multiple choice format. Thus, for the participants in this study, the multiple choice recognition task provided additional information about memory functioning over and above that obtained from the standard administration of the Visual Reproduction subtest.

Perceptual Matching

Following multiple-choice recognition, each participant was given a perceptual matching task regardless of their memory recall. Ninety-three percent of the control group had 100% accuracy on this task, indicating that less than perfect matching was unusual in a control group. Sixty percent of the experimental group had 100% accuracy on perceptual matching. Of the 40% (12 participants) who had less than perfect matching, 75% were able to match 4 out of 5 of the designs.

Further investigation of the 12 experimental participants who had less than perfect matching revealed that: 1) four had no delayed or cued recall of the design/s that they were unable to correctly match, 2) three had no delayed recall but were able to recall information when provided with cues, and 3) the remainder had only minimal recall of the designs (a score of 10 or less).

The three experimental participants who matched less than three of the designs, had minimal or no delayed recall of these designs, and also performed poorly on multiple choice recognition.

Taken together, the results from the experimental group suggest that when participants are unable to correctly match a design, their recall is minimal or zero. Hence, their perceptual difficulties appeared to be a significant factor in their defective recall.

Reliability

The internal consistency of multiple choice and perceptual match tasks was examined by computing Cronbach's alpha. This was computed using the combined data from the control and experimental groups (n=60). The multiple choice task had an alpha reliability of .79, whilst the perceptual match task had an alpha reliability of .33. Thus, the internal consistency of the multiple choice task was satisfactory, whilst the reliability of the perceptual match task was questionable on the basis of Cronbach's alpha. The small range of items and the reduced range in the scores obtained are likely to be the basis of the lower reliability of the perceptual match task.

Validity

In addition to examining the validity of the revised scoring system for the Visual Reproduction subtest, the concurrent and construct validity of the new measures included in the revised Visual Reproduction scoring system, namely the multiple choice and perceptual match tasks, were examined. However, the poor reliability of the perceptual match task is likely to impact on its validity and therefore caution needs to be taken when evaluating these results.

Criterion Related Validity

Criterion related validity was examined by investigating the relationship between the standard delayed recall administration of the Visual Reproduction subtest and the additional measures designed in this study. Table 63 illustrates Spearman's correlation between memory for designs and the multiple choice and perceptual match measures for the control and experimental subjects respectively.

Table 63

Correlation between Visual Reproduction Immediate and Delayed recall with the Multiple Choice and Perceptual Matching tasks

	Control		Experimental	
Total Score	Multiple Choice	Perceptual Match	Multiple Choice	Perceptual Match
Immediate				
Revised	.33	.10	.59	.65
Original	.35	-.10	.55	.61
Delayed Recall				
Revised	.40	.03	.66	.33
Original	.37	.10	.68	.43

When examining the relationship between memory for designs and the multiple choice task, there was a moderate correlation in the control group and a large correlation in the experimental group. These results suggest that there was a moderate to strong relationship between better recall of the designs and recalling more items when given a recognition format.

In regard to the perceptual match task, there was only a weak correlation with immediate and delayed memory for designs using the revised and original scoring system in the control group.

In the experimental group, the relationship between the Perceptual Match task and memory for visual reproductions was generally stronger than in the control group and was in the moderate to large range. This result suggests that performance on free recall is at least moderately related with performance on the perceptual match task, with a better level of free recall being related to a better performance on the perceptual match task.

Construct Validity

In order to examine construct validity, the performances by the experimental and control groups on the multiple choice task and perceptual match task were compared.

Table 64 illustrates the means and standard deviations for the Visual Reproduction multiple choice and perceptual match measures designed in this study. The scores on these two tasks were positively skewed and therefore nonparametric statistics were used to compare the performance by the two groups.

Table 64

Mean Scores on the Multiple Choice and Perceptual Match tasks

	n	Mean	Standard Deviation
Multiple Choice			
Total	60	3.57	.98
Experimental	30	3.27	1.2
Control	30	3.87	.57
Perceptual Match			
Total	60	4.70	.62
Experimental	30	4.47	.78
Control	30	4.93	.25

Mann Whitney U comparisons between the control and experimental group in these measures is presented in Table 65. The Visual Reproduction multiple choice and perceptual match trials significantly discriminated between the two groups. The experimental groups performed more poorly on both of these tasks.

Table 65

Mann-U-Whitney Test Comparisons Between the Experimental and Control Groups on the Multiple Choice and Perceptual Match Task

	z	df	Probability
VR Multiple Choice	-2.562	58	.010
VR Perceptual Match	-3.069	58	.002

DISCUSSION

The Visual Reproduction subtest, which has been part of the Wechsler Memory Scale since its conception, has been widely used as test of memory in clinical practice and in research studies (Bigler et al., 1996; Chelune & Bornstein, 1988; Fastanau, 1996; Haut et al., 1994, 1996; Larrabee & Chelune, 1995). This study sought to develop a revised scoring system for the Visual Reproduction Subtest of the WMS-R and then explore a specific measure of non-verbal memory based on that revised system. In addition, an elaborated administration involving cued recall, recognition testing and perceptual match measures was developed and trialled.

Part One- Evaluation of the Revised Scoring System

Design Advantages of the Revised Scoring System

A revised scoring system for Visual Reproduction subtest of the Wechsler Memory Scale-Revised was developed in this study to address the problems and/or limitations identified with the original scoring system. However, the revised scoring system was not intended to be a radical departure from the original, rather it was conceptualised as a scoring system that contained fewer anomalies but still reflected a similar grading of memory.

A number of limitations have been identified in the design of the original scoring system. For example, when using the original scoring system, not all items are applied independently and so the contribution of a key item to the overall score may be inflated and some recall may not be rewarded/acknowledged. Hence, the revised scoring system was designed so that each item could be scored independently. In this way the revised scoring system makes few qualitative assumptions about the way information is retained and just grades the amount of memory recall.

Items in the original scoring system did not include allowances for carelessness in drawing, poor drawing ability or reduced motor functioning and also tended to focus on measurement of precise angles and distances. In the revised scoring system, items were developed so that they included generous tolerances.

Thus, the occasional lapse (for e.g. forgetting one dot) or poor motor control or drawing ability (for e.g. having gaps between the join of lines and overshoots across lines) was not penalised. Items also included generous tolerances for the measurement of angle degrees and the length of lines, reducing the demands on subjective judgement, making scoring less dependent on precise drawing and returning the focus of scoring to memory recall.

The revised scoring system developed in this study also included a number of items that addressed incomplete or partial recall. Although participants are often noted to recall partial elements of a design in clinical practice, such recall may not receive any credit on the original scoring system. This is a potential limitation because a score could underestimate the extent of recall. As a result of including partial credit, the revised scoring system has the potential of being more able to adequately represent the extent of memory recall and address different recall quality.

The number of items for each of the designs on the WMS-R original scoring system was not equal and there was no rationale or logical basis presented for this differential weighting provided in the WMS-R manual. There was also a limited range of items for the first three designs. The revised scoring system was devised so that an equal number of points were allocated to each of the four designs. The inclusion of more items increased the range of scores available and the potential for discriminating between different quality/quantity of recall.

A major criticism of the Visual Reproduction subtest has been that the stimuli are potentially verbally encoded (Heilbronner, 1992; Lee et al., 1989). The revised scoring system was designed to include items that placed a greater emphasis on aspects of the designs that are non-verbal or less likely to be verbally encoded, such as the spatial relationships between elements. Nevertheless, the potential for verbal encoding of a substantial number of aspects remains.

Psychometric Properties of the Revised Scoring System

In order for the revised scoring system to be clinically useful it has to be reliable. Moreover, to be considered as a real alternative to the original scoring system, the revised scoring system must have equal or greater reliability than the original scoring system. The results in this study indicated that there was a very high degree of consistency between the scores obtained on two occasions by a single rater (intra-rater reliability) when using the revised scoring system. This high level of consistency was marginally higher than that found for the original scoring system. Thus, the interpretation and application of the revised scoring system was stable over time. There was also a very high level of consistency between the scores on each design obtained by two separate raters using the revised scoring system (inter-rater reliability). The two raters in this study appeared to apply the items in the revised scoring system with a high level of consistency.

The intra-rater and inter-rater reliability of the original scoring system was also found to be very high in this study. Two independent raters applied the original scoring system in a consistent fashion, producing a high level of agreement in scoring. The reliability of the original scoring system was very similar to that obtained for the revised scoring system with the same two raters. However, in view of the increased number of items on the revised scoring system (80 items vs 41 items) there was more potential for variability on the revised scoring system. Given the greater number of items on the revised scoring system, the similar degree of reliability suggests that the revised scoring system has marginally better scoring reliability. However, it is pertinent to note that the original scoring system can have high inter-rater reliability when applied rigorously.

The internal consistency of the revised scoring system in this study sample was also quite high when examining immediate and delayed recall. The internal consistency was higher than for the WMS-R original scoring system standardisation sample, the WMS-111 version of the task and the original scoring system in this study.

The inter-rater reliability of the original scoring system has been examined by a number of authors (McGuire & Batchelor, 1998; Wechsler, 1987; Woloszyn et al., 1993). The results from the combined control and mixed clinical sample in this study were similar to that of the Woloszyn et al. (1993) mixed clinical sample and Wechsler's (1987) normal sample. Wechsler (1987) reported inter-scorer reliability coefficients of .97 in the original standardisation sample of the WMS-R and Woloszyn et al. (1993) reported reliability coefficients of .977 for immediate recall and .975 for delayed recall in a clinical population. These results are consistent with the findings of the current study.

When examining inter-rater reliability in a clinical population, McGuire and Batchelor (1998) reported Pearson's coefficients of .82 for immediate recall and .94

delayed recall, although reliability coefficients on individual Designs ranged from .49 to .94. Although the coefficients for total recall were only marginally different from those found in the current study, differences in coefficients on individual designs varied more widely between these studies and were consistently smaller in the McGuire and Batchelor (1998) study. However, the mixed clinical population in the current study was quite different to the population in the McGuire and Batchelor (1998) study; namely individuals who had undergone temporal lobectomies. It is possible that the reliability of scoring varies across certain clinical groups. Hence, further evaluation of the scoring of visual reproductions in different clinical populations may be warranted.

A number of studies have found high levels of inter-rater reliability on the Visual Reproduction subtests. However, differences between the raw scores obtained on individual designs across raters have been found to be large in some cases. For example, Wechsler (1987) reported that the raw score on each design differed by up to four points across raters, however large differences were likely to have been infrequent given that Wechsler (1987) found that the average raw score difference in total between scorers was 1.5 points. Similarly, McGuire and Batchelor (1998) also reported that raters varied by up to four points on each design. Again, the overall mean difference between scorers was only four points, suggesting that very large differences may have been uncommon. Five of their forty participants had total scores that differed by 10 points or more between the two scorers, suggesting that large differences were infrequent.

A substantial difference in scores between the two raters was a rare phenomenon in this study, when using either the original or the revised scoring systems. Although the results of this study showed that the difference in scoring on each individual design between two scorers could potentially be quite high (for e.g. 6 points on Design 3 when scoring immediate recall), the frequency of a substantial difference was quite rare. The difference between the mean score of each rater was very small across designs, indicating that generally the raters were obtaining similar scores.

Overall, the results of this study provided support for a high consistency in scoring with both systems, although there appeared to be a marginal gain in reliability when using the revised scoring system. The very high consistency in scoring with the revised scoring system supported the overall aim of developing the revised scoring system and the principles used in generating the items. When using the original scoring system in clinical practice, there is a potential for individual items/criteria not to be strictly applied because the attention to detail of these items is not seen as reflecting memory per se. Developing more explicit criteria to improve scoring reliability, including specifying more generous tolerances in measurement on a number of items and increasing the allowances for carelessness in drawing or poor visuo-motor skills, were seen as having the potential to improve scoring reliability. Thus, the finding of very high reliability in scoring, both across scoring occasions and across scorers is consistent with the overall aim of developing a revised scoring system.

Increased Score Range

One of the major design features of the revised scoring system was that it contained more items than the original. The original scoring system had a limited number of items (41 compared to 80). This is particularly true for the first three designs on the WMS-R that collectively contributed only 54% to the total score on the original scoring system. These same items contribute 75% to the total score on the revised scoring system. The small range in scores on the original scoring system has the potential to reduce the discriminatory power of the subtest. Thus, increasing the number of items on the revised scoring system to 20 items for each design potentially increases the possible range of scores obtainable on this task, particularly for the first three designs.

The possibility of an increased range does not necessarily mean that the scores found in studies will have a greater range. When examining the range of scores obtained in the combined study group (experimental and control participants), there was a larger range of scores generated by the revised scoring system for each of the designs and for total recall on both immediate recall (Revised 54 versus Original 33) and delayed recall (Revised 75 versus Original 39). Whilst the range on the original scoring system was generally as wide as the number of scoring items allowed (e.g. 7 for Design 1), the range on the revised scoring system was larger.

The interquartile ranges found when examining the data for the total sample were also wider for the revised scoring system, particularly when examining delayed recall (Revised 32.25 versus Original 16 for the total score on delayed recall). This was also true when examining the control data in isolation (Revised 25 versus Original 16.5 for the total score on delayed recall). Thus, the middle 50% of scores fell within a wider range on the revised scoring system. Since delayed recall is considered to be the most useful indicator of memory functioning, a wider dispersion of scores potentially allows for better discrimination of differences in memory performance. In addition, a wider score range allows more scope for detection of changes over time.

Increasing the number of items for each design also resulted in there being a larger range of scores below the 25th quartile when scoring delayed recall with the revised scoring system (Revised 23.5 versus Original 9 for the total score on delayed recall). It is this area where it is important to have a substantial score range so that floor effects are minimised in impaired populations.

Although the increased range of scores found when using the revised scoring system could potentially provide better discriminatory power at higher and lower levels of performance, further research will need to be conducted in order to examine this.

Relationship between the Original Scoring System and the Revised Scoring System

The revised scoring system was designed to contain few anomalies than the original whilst generating a similar grading of memory. Thus, it was not intended that the revised system was radically different from the original scoring system. The results in this study suggested that the two scoring systems provide a similar grading of memory functioning when applied rigorously. The correlation between systems was quite high supporting the principles used in developing the revised scoring system. There was also a linear relationship found between the scores obtained on the original and revised scoring system for the specific sample in this study. This holds promise for developing a conversion equation to convert scores on the revised system to scores on the original system.

The very strong agreement association between the original and revised scoring systems means that the revised system could replace the original system without any major change in the grading of memory performance. This makes the transition to using the revised scoring system less problematic, because using the revised scoring system to rate visual reproductions would not entail losing any benefits on the old system.

The very strong relationship between the two scoring systems was not an artifact of the scoring systems having a high number of identical items. In fact there were only 10 identical items on the two scoring systems (25% of the original items, 12.5% of the revised items). Moreover, the fact that the scorers were aware of the purpose of scoring would have not influenced the degree of the relationship between the two systems per se, although it could well have affected the stringency with which the two raters applied the scoring systems. What the results have indicated is that when the two scoring systems are applied rigorously, the result is a strong relationship between them. This is consistent with the aim of the study, which was to develop a revised scoring system that addressed the limitations of the original, but still provided a similar grading of memory functioning.

Moreover, use of the revised scoring system does not require any alteration to the standard administration of the Visual Reproduction subtest. Thus, the revised scoring system could be used as an adjunct to the original. This could offer an advantage in clinical situations where a clinician considered that a score on the original scoring system underestimated recall. Since the two scoring systems are highly related and there appears to be a linear relationship between performance on the revised scoring system and the original scoring system, it would be expected that there would be a high relationship between the two scores. If a significant discrepancy was found in favour of the revised system this would lend support to the clinical impression that the original system underestimated memory function in that individual.

Disadvantages of the Revised Scoring System

There are a number of potential disadvantages with using the revised scoring system. Firstly, there is no normative data about performance on the revised scoring system at the present time. However, given the linear relationship found between the two scoring systems in this study, a score conversion from the revised to the original scoring system may be possible so that the normative data for the WMS-R could be used. This would not be highly problematic because there was a high correlation between the two scoring systems.

The strong linear relationship found in this study was based on a relatively small sample. Therefore to confirm that a score conversion from the revised system to the original scoring system is appropriate, larger samples would need to be assessed. Although it would be possible to use a score conversion in order to evaluate performance on the revised scoring system, the WMS-R norms have been extensively criticised. Therefore a normative study using the revised scoring system would be more appropriate and clinically relevant.

A further disadvantage in using the revised scoring system is that data on the scoring reliability in a clinical setting is not available. Although the results in this study illustrated a high level of agreement between the two raters, it is important to note that the two raters in this study were involved in the development of the revised scoring system. Thus, they were not applying the items directly from printed criteria, but rather after considerable exposure to the individual items and their interpretation. These raters also had considerable practice in applying the revised scoring system on a variety of designs. This may have resulted in a higher level of consistency in scoring than what would be found in scorer's naive to the development process. However, the explicit content of the items of the revised system would suggest that it is unlikely to be less reliable than scoring with the original system. Nevertheless, it would be important to confirm this with raters who only learned the criteria directly from the printed criteria. Further research could examine the accuracy of applying the revised scoring system with new users.

The two raters in this study were also aware that they were rating the designs for a research study and they scored a number of protocols in the same sitting. Therefore, it cannot be assumed that the reliability will be as high in clinical practice when protocols are typically scored individually. However, this study has shown that when the scoring items are applied strictly, a high level of consistency in scoring can be obtained.

With the 50% increase in the number of items on the revised scoring system, the time taken to score designs with the revised system is highly likely to be longer than the time taken to score the original system. This is likely to be more prominent with clinicians who are experienced in applying the WMS-R criteria. However, with new clinicians the difference may be minor because the revised criteria requires less judgement.

Although there is currently little normative information available and it may take more time to score, the revised scoring system has considerable potential to assist in understanding memory functioning in clinical practice. It is very reliable, generates a wide range of scores and produces a similar grading of memory functioning to the original scoring system. Further studies involving the development of a normative base and collection of data in the field would be advantageous.

Part Two- Assessment of Non-Verbal Memory with the Revised Scoring System

A wealth of research has explored what specific aspects of memory function are assessed by the Visual Reproduction subtest, mindful that it was included in the WMS-R as a contrast to the auditory-verbal memory tasks (Chelune & Bornstein, 1988; Fastanau, 1996; Haut et al., 1994, 1996; Larrabee & Chelune, 1995; Loring & Papanicolaou, 1987). Although the Visual Reproduction subtest has often been used as a measure of non-verbal memory, support for this assumption has been inconclusive. For example, the subtest has been criticised as not being a particularly valid assessment of non-verbal memory because the stimuli could potentially be verbally encoded by some individuals (Heilbronner, 1992; Lee et al., 1989). Indeed, studies have shown that there is a substantial correlation between performance on this subtest and on verbally mediated tasks (Chelune & Bornstein, 1988; Larrabee & Chelune, 1995). Additionally, performance on the subtest does not correlate reliably with right hemisphere damage which would be predicted by the material specific model of memory (Bigler et al., 1996; Chelune & Bornstein, 1988; Naugle et al., 1993). Furthermore, there appears to be a strong association with non-memory factors, as indicated by the substantial loadings on visuo-constructional and visual-perceptual factors in factor analytic research (Larrabee & Curtis, 1995; Leonberger et al., 1991). Indeed, a major criticism of the Visual Reproduction subtest has been the potential for scoring and interpretation on this task to be confounded by impairments in constructional, perceptual and/or motor skills (Gfeller et al., 1995; Haut et al., 1994; Ricker et al., 1994).

One of the design features of the revised scoring system was the theoretical inclusion of more items that were potentially non-verbally encoded. Thus, the revised scoring system could offer a better measure of non-verbal memory functioning than the original system. Given the inconclusive support found in other studies for the original system, it was hoped that the inclusion of more "non-verbal" items would improve the subtests ability to reflect non-verbal memory.

Relationship between the Revised Scoring System and Non-Verbal and Verbal Memory Measures

In order to provide support for the Visual Reproduction subtest as a measure of non-verbal memory when using the revised scoring system, the performances of the control and experimental groups were compared with the results on a number of other memory tasks. A strong association with other tasks thought to tap non-verbal memory processes would provide indirect support for the Visual Reproduction subtest being a measure of non-verbal memory.

The results in this study appeared to support the idea that in an unimpaired population, memory for designs shared a stronger relationship with memory for faces, than with verbal memory tasks that have a strong language basis. There was a moderate association between performance on the Visual Reproduction subtest (scored on the revised scoring system) and the WMS-111 Faces subtest in the control group. This relationship was stronger when examining delayed recall.

The moderate to larger relationship between performance on these two subtests was particularly favourable given that one task involves a recognition format (Faces) and the other involves free recall (Visual Reproduction). The degree of association suggests that these two tasks measure at least some common aspect of memory functioning, potentially non-verbal memory. Past research that has shown an association between the right hemisphere of the brain and performance on face memory tasks, would support the notion that both tasks draw on non-verbal memory (Morris et al., 1995b; Naugle et al., Herman et al., 1995a; Warrington, 1984). Thus, the relationship between memory for designs and memory for faces is hypothesised to be based on their reliance upon primarily non-verbal rather than verbal memory process, although no doubt verbal memory processes play some role as well.

In the control group, the Logical Memory subtest shared a moderate relationship with the Visual Reproduction subtest when examining immediate recall. The relationship between memory for designs and memory for stories was reduced when examining delayed recall. Given that delayed recall is likely to represent a better measure of memory, this suggests that the recall of designs may be less influenced by verbal processes when delayed recall is examined. Based on these results, the Visual Reproduction subtest did appear to share a marginally stronger relationship with the Faces subtest than with the Logical Memory subtest in the control group, particularly on delayed recall. This is consistent with the idea that the relationship between non-verbal measures was stronger than the relationship between verbal and non-verbal measures.

The correlation between the Logical Memory and Visual Reproduction subtests underscores that there is some shared variance between them. Indeed, memory factors in factor analytic studies generally have loadings from both verbal and non-verbal memory tests particularly with increasing age (Larrabee & Curtiss, 1995; Larrabee et al, 1985; Smith et al, 1992a). This suggests that both verbal and non-verbal memory procedures share a degree of common variance in most groups. The relationship between the Vocabulary subtest and the Visual Reproduction in this study was also moderate to large in the control and experimental groups. These results are consistent with the idea that language plays a role in memory for non-verbal material, although the influence of language may be different across individuals and groups.

There would appear to have been at least a moderate relationship between Visual Reproduction and the recall of word pairs in the control group. The immediate recall of associated word pairs generally had a moderate correlation with memory for designs. The magnitude of this relationship was reduced when examining delayed recall, particularly when examining the revised scoring system.

The association between memory for designs and "novel" pairs was marginally stronger, with this relationship being similar in magnitude to that found between the Visual Reproduction and Faces subtests, two non-verbal tests. This finding is not unique to this study. Trahan et al. (1988) found a moderate correlation between delayed recall on the Visual Reproduction subtest and the Expanded Paired Associates Test. These findings suggest that the performance of individuals on the Visual Reproduction subtest may be dependent on both non-verbal and verbal encoding strategies. They could also suggest that verbal material can be processed non-verbally. Trahan et al (1988) reported that a number of normal participants spontaneously reported using visual imagery to help them learn word pairs. Interestingly, the Faces subtest also shared a moderate relationship with recall of "novel" word pairs, again suggesting an association between non-verbal and verbal processing of material regardless of the mode of presentation. It is also possible that the recall of "novel" pairs depends on the learning of arbitrary associations (Saling, 1998), as does recall of new visual material. The basis of this relationship would appear to warrant further investigation.

Interestingly, recall of "novel" word pairs was also shown to share a strong relationship with the block design subtest of the WAIS-R (a measure of visuo-constructional ability) in both the control and experimental groups. Similarly, the Visual Reproduction subtest shared a very high correlation with Block Design. This provides indirect support for the idea that similar processes are involved in memory for "novel" word pairs and designs.

The correlation between the verbal memory measures used in this study, Logical Memory and Verbal Paired Associates, was generally small in the control group, although it would be expected that these measures would share a high correlation. This is not consistent with the moderate to large correlation found in the WMS-R standardisation sample, particularly in the older age groups.

It is possible that the low correlation between recall of word pairs and prose passages in the current study occurred because the "novel" and associated pairs were separately compared with Logical Memory. Thus, in the WMS-R standardisation sample, combining "novel" pairs, a language poor task, with associated pairs, a language rich task may have increased the relationship between the word pairs and stories. However, in this study recall of "associated pairs", a language rich task, also did not have a strong correlation with the Logical Memory subtest, another language rich task. This finding is unexpected. It is possible that different sample characteristics between the WMS-R standardisation study and this study contributed to this difference, as to could the inclusion of an additional two "associated" word pairs (north-south, up-down).

Direct comparison of the relationships between memory measures in this study with those reported in the WMS-R and WMS-111 are difficult due to differences in the composition of the Logical Memory, Visual Reproduction, and Verbal Paired Associates subtests across the WMS-R, the WMS-111 and this study. Thus, the comparisons made here need to be treated cautiously. When examining the relationship between the Visual Reproduction and Logical Memory subtests in the WMS-R standardisation sample, there was generally a weak relationship between these measures in the two younger and the oldest age group, but moderate to large relationship in the 35-69 age groups for both immediate and delayed recall. Given that the sample used in this study were of an older age, these two studies are generally consistent in the trend for memory measures (regardless of content) to share a higher association with each other as age increases. Bornstein and Chelune (1989) also found that there was a greater loading of non-verbal memory tests on a verbal memory factor with increasing age in factor analytic research. This again suggests that the association between memory measures may strengthen with age.

The correlation found between memory for designs and recall of associated and "novel" word pairs in this study was generally larger than the relationship between memory for designs and memory for word pairs found in the WMS-R standardisation sample, particularly when examining the revised scoring of the Visual Reproduction subtest. However, the recall of familiar and novel pairs were combined into one score on the WMS-R, whereas they were examined separately in this study. This could suggest that combining the familiar and novel pairs in one score masks some variability in the relationship with other measures.

The moderate relationship between memory for designs and memory for faces found in this study was larger than the weak relationship found between the Visual Reproduction subtest and the Faces subtest in the WMS-111 standardisation sample. The difference in the strength of the relationship cannot be attributed to the Faces subtest as it is the same task in both studies. The difference could be an artifact of the different content and scoring of the WMS-111 Visual Reproduction subtest as compared to the WMS-R version. As has been outlined previously, there are a number of potential limitations with the WMS-111 version of the task and this may result in the correlation reported in the WMS-111 underestimating the relationship between these two tasks.

In the experimental group, there was little difference in the strength of the relationships between modality specific memory tasks. When examining the relationship between performance on the memory tasks, there was generally a moderate to large relationship between all measures, particularly when examining delayed recall. This could be interpreted as suggesting that a general memory factor is more prominent in individuals with brain impairment. However, the experimental group was a mixed sample, with a variety of neurological conditions. It may be that groups with circumscribed lesions in the left and right hemispheres would demonstrate a stronger relationship between modality specific tasks than between tasks assessing different aspects of memory.

The results in this study suggest that there is a difference in the relationship between memory tasks in impaired and unimpaired populations. In understanding the performance of the experimental group, it may be that a general memory factor is more prominent. In individuals with brain injury, normal divisions between processing different types of material may be disrupted and therefore these indivduals may rely on a combination of processes, including both verbal and non-verbal encoding of material, in order to remember information. In contrast, the control group is only experiencing normal age changes and may continue to rely on different processes in order to perform different types of memory tasks. It is possible that for individuals without brain injury, non-verbal and verbal memory tasks share some degree of shared processing (indicated by the absence of a zero correlation), but also a large degree of distinct processing.

The Ability of the Revised Scoring System to Distinguish between Clinical Meaningful Groups

Lateralised cerebral damage has been found to result in modality specific memory impairment, with the relationship between the left hemisphere and verbal memory impairment being well established (Chelune et al, 1991; Herman et al., 1995a; Saling et al., 1993). The converse association, that between the right hemisphere and non-verbal memory, although well accepted clinically, has not been consistently established (Hermann et al., 1995a; Rausch & Babb, 1993; McMillan et al, 1987; Sass et al., 1992). Specific evidence for an association between the Visual Reproduction subtest and the right hemisphere has also been inconsistent (Bornstein et al., 1988; Delaney et al., 1980; Naugle et al., 1993).

The ability of the Visual Reproduction subtest to discriminate between groups with left and right hemisphere brain damage was also investigated in order to provide support for the non-verbal nature of the subtest. It was expected that scores on the Visual Reproduction subtest would be different between two groups with clearly lateralised brain damage, because of the association between non-verbal memory and the right hemispheres of the brain.

As was expected, the group with right hemisphere damage in this study performed significantly more poorly than did the left hemisphere group on the Visual Reproduction subtest. This finding was consistent on both scoring systems when examining the total score for immediate and delayed recall. Since non-verbal memory is considered to be a function of the right hemisphere, the poorer performance of the right hemisphere group on the Visual Reproduction subtest supports the construct validity of the Visual Reproduction subtest as a test of non-verbal memory.

In addition to performance on the Visual Reproduction subtest being significantly different between the groups with left and right hemisphere damage, the scoring systems had a strong ability to discriminate between the two groups. In this study, the original scoring system classified more of the right hemisphere group correctly than the revised scoring system (15 on the original system as compared to 14 on the revised system). The revised scoring system classified more of the left hemisphere group correctly (14 on the revised system as compared to 13 on the original system). Thus, both scoring systems appeared to have good sensitivity and the difference between the two systems in classification rate is marginal. These results suggest that Visual Reproduction as scored on both scoring systems, has a very good ability to distinguish between those who have a non-verbal memory difficulty and those who have not. However, the sensitivity and specificity of the Visual Reproduction subtest as scored on both systems would need to be replicated, given that the data in this study, due to skewness and kurtosis, violated the assumptions underlying parametric statistics.

Previous research has also found that the Visual Reproduction subtest has utility in distinguishing left and right hemisphere brain injured groups in some samples. For example, Delaney et al. (1980) found that a right temporal lobe epilepsy group were impaired on the Visual Reproduction subtest as compared with a left temporal lobe epilepsy group. However, these groups differed only when percent retention was used as the measure of memory, rather than the total score on immediate and delayed recall. Bornstein et al. (1988) also did not find that a right temporal lobectomy group performed significantly worse on the Visual Reproduction task. However, they did report that participants with right hemisphere damage were more adept at learning verbal than non-verbal material. Conversely, participants with left hemisphere lesions displayed better learning and retaining nonverbal/visual material than comparable verbal material. Thus, the two groups did differ in terms of their patterns of performance across modality-specific tasks.

In 1993, Naugle et al. found that a group with left hemisphere damage had a significant discrepancy between their verbal and non-verbal memory performance on the Logical Memory and Visual Reproduction subtests of the WMS-R. However, no such discrepancy was found for the group with right hemisphere compromise. Sass et al. (1992) found that immediate recall of Visual Reproduction subtest of the WMS distinguished between groups with left and right temporal lobe epilepsy but delayed recall and percent retention did not. However, no measure correlated with the extent of hippocampal neuronal density.

Although the current study found that the Visual Reproduction subtest distinguished between groups with left and right cerebral compromise, as has been illustrated this is not a consistent finding. The difference between these findings could relate to different sample characteristics. For example, there is a great heterogeneity in outcome following temporal lobectomy and therefore it is possible that combining a number of participants into groups could mask any association between the Visual Reproduction subtest and poorer outcome after damage to the right temporal lobe. (Baxendale, 1997; Heilbronner et al., 1989; Heilbronner, 1992; Hermann et al., 1992; Lee et al., 1989; Powell et al., 1985). As has been found in this study, there may be a stronger association between memory tasks with increasing age. Thus, the lack of an association between the Visual Reproduction subtest and right temporal damage may be a result of the strong association between the verbal and non-verbal memory tasks in the samples used in some studies.

The results of this study may have been consistent with the material specific memory model because of the characteristics of the participants in the left and right hemisphere groups. The right and left hemisphere groups in this study were a heterogenous group who had damage to the right hemisphere as the result of stroke. Therefore, it is possible that damage was not only restricted to the mesial temporal region. The studies outlined previously were generally conducted with groups with damage combined to the mesial temporal region. It could be that the Visual Reproduction subtest is sensitive to distributed damage rather than damage confined to the right mesial temporal region.

In summary, the results of this study supported that the Visual Reproduction subtest, as scored by both the original and the revised scoring system, was able to discriminate between clinically meaningful groups. The association between poorer performance on the Visual Reproduction subtest and right unilateral hemisphere damage provides indirect support for the subtest being a measure of non-verbal memory functioning.

A Non-Verbal Memory Index

Although tasks can be described as measuring non-verbal memory, it is likely that people use verbal skills when remembering non-verbal information. The association found between verbal and visual memory measures on factor analysis supports this (Larrabee & Curtiss, 1995; Larrabee et al, 1985; Smith et al, 1992a). In addition, because many "non-verbal" memory tasks, such as the Visual Reproduction use simple geometric figures, they are likely to be verbally encoded by some participants. Although new tests have been designed using complex and unfamiliar stimuli, these may still be verbalisable to some extent (Eadie & Shum, 1995; Vanderplass & Garvin, 1959). An alternative to using complex figures is to acknowledge that non-verbal memory tasks are likely to be verbally encoded at least to some extent by some individuals. Thus, memory for non-verbal material is likely to be the result of a combination of processes and some amount of verbal mediation may take place. Indeed, persons with left cerebral damage can exhibit impairments on this task (Trahan et al, 1988). The poor performance on this task by individuals with left hemisphere damage may be because of deficits in verbal encoding of the material.

Although both verbal and non-verbal processes may contribute to a person's memory of non-verbal material, a method for extracting non-verbal memory processing from tasks that can be encoded in both ways would be useful clinically. This study aimed to produce an Index to identify deficits in non-verbal memory functioning. Rather than eliminating the potential for verbal encoding by altering the complexity of the stimuli, an Index was derived from the items on the Visual Reproduction subtest that were more likely to be addressing the non-verbal aspects of the material. Thus, the Index developed extracted those items that were potentially more non-verbal memory to be separated from those aspects that were potentially more verbal. In this way, the Index derived could be used to identify non-verbal memory difficulties. This approach acknowledged that some scoring items probably reflects verbal encoding, whilst other items might be better at addressing aspects of non-verbal memory.

The Index developed in this study was not based on any prior model or division of function, it was simply extracted from clinical data. Each item that contributed to the Index was derived entirely from the data obtained in this study. Since the Visual Reproduction subtest was found to distinguish between left and right hemisphere groups, the scale was devised of items (rather than entire designs) that were performed significantly more poorly by the right hemisphere group. It was considered that these items were more likely to reflect non-verbal memory processes. However, when finalising the composition of the Index, a number of exclusionary criteria were used. Firstly, only items that discriminated between groups on delayed recall were included because delayed recall is considered to provide a better indication of memory function. Secondly, any item that was found to be significantly different between the two groups on both immediate and delayed recall was also excluded. This exclusion was conducted in order to eliminate or minimise other confounding deficits like visuo-construction abilities. If an item was performed poorly by both groups, it was also excluded, regardless of a significantly different performance. This was considered necessary so the scale was sensitive to non-verbal memory impairment, rather than memory impairment in general. The global rationale of these exclusions was to develop a scale that was more likely to be tapping memory functions rather than confounding factors such as other like non-verbal cognitive abilities.

A 16 item Index of non-verbal memory functioning was identified from the revised scoring system. In order to evaluate this potential Index a linear discriminant function was conducted. This group of 16 items was able to correctly classify all the right and left hemisphere patients.

The Index derived from the revised scoring system had very good sensitivity in distinguishing those with left and right cerebral lesions. However, the Index was derived from the same data as it was validated against (the left lesion and right lesion group). Thus, the discriminating power of the Index was likely to have been an overestimate. In order to provide more support for the Index, it will be important for future research to examine its ability to discriminate between clinically meaningful groups. If it proves to have good sensitivity in these groups, normative data will then need to be collected for scores on this Index. In addition, to provide further support for the non-verbal nature of the Index, further research could examine the relationship between the Index and other tasks thought to measure non-verbal memory. Unfortunately, this data was not available for all members of the left lesion and right lesion groups in this study.

It is also possible that the discriminatory power of the scale in this study was an artifact of factors other than memory. For example, the difference between the groups on individual items could be related to differences in general intelligence or visuo constructional ability. Information about cognitive functioning and non-verbal cognitive functioning in particular was not available for all the participants in the left lesion and right lesion groups. Further research will need to be conducted to demonstrate that the Index reflects non-verbal memory processes, rather than non-verbal or general cognitive ability.

The Index derived in this study was based on a very small number of participants. It is therefore possible that the Index composition may vary with data from a larger number of participants. However, the concept of a non-verbal memory Index has considerable clinical utility. By using such an Index, a non-verbal memory difficulty could be identified without changing the content of the task. The principal behind devising an Index is that future identification of non-verbal memory deficits should focus on lateralisation by items rather than lateralisation by overall scores. This approach recognises that some parts of current tests may be better measures of non-verbal memory.

In summary, a potentially useful Index of non-verbal memory functioning was derived from the revised scoring system developed in this study. This Index was able to identify those with right hemisphere lesions with good sensitivity. Although the Index may be potentially useful in identifying individuals with non-verbal memory deficits, it does not provide information about the severity of the non-verbal deficit so far. Further work will need to establish if this Index has sensitivity and specificity in identifying people with non-verbal memory impairment in other groups and whether it quantifies the extent of the deficit.

Part Three- Expanding the Clinical Utility of the Visual Reproduction Subtest

When assessing memory, it is important to include tasks that provide information about how well the participant has encoded, stored and retrieved information. Clearly, a failure to recall information in a free recall procedure does not, at least in isolation, imply that the information was not encoded or that it has been forgotten. Instead, a failure to recall information could result from a difficulty in retrieving material that had been stored in memory. Thus, information needs to be obtained from assessment procedures in order to make an analysis of the extent information is acquired and stored and the effectiveness of retrieval processes in accessing information. This is diagnostically important as some memory disorders involve a prominent encoding deficit (Alzheimer's Disease and Wernicke Korsakoff Syndrome) whilst others involve a retrieval deficit (Huntington's Disease) (Lezak, 1995). It also has rehabilitation implications, as those with a prominent retrieval deficit are likely to benefit from cueing, whilst those with an encoding deficit are not.

Traditionally, assessment tools have relied on free recall procedures, however using free recall alone fails to reveal the basis of impaired performance. In order to make inferences about the basis of a memory impairment, assessment tools need to include free recall and recognition procedures. A high score on the free recall portion of a task is likely to indicate effective encoding and retrieval of the material. A high score on the recognition portion of a task is likely to be indicative of successful encoding and storage of the material, regardless of performance on free recall.

The WMS-R version of the Visual Reproduction subtest did not include a measure of recognition and therefore an evaluation of the contribution of encoding, storage and retrieval to memory performance could not be made. Although the WMS-111 version included additional measures to assist in quantifying the contribution of storage and retrieval to performance there were a number of problems with their design and administration. In view of the problems with the WMS-111 recognition measure, standardised cued and recognition procedures were developed in this study to assess the contribution of effective encoding and retrieval to performance on the Visual Reproduction subtest.

Cueing – The Contribution of Encoding and Retrieval to Memory Recall

In order to partial out the contribution of a retrieval deficit to memory recall, clinicians have seen value in using a recognition task to ascertain if material has been stored. However, an intermediate step between free recall and recognition is the use of a cue to facilitate recall. Cued recall differs from recognition in that not all the stimulus is presented. The rationale for a partial cue, rather than the presentation of the whole stimuli, is that it may be sufficient to trigger recall in some people, without the need for presenting the complete item in its original form.

The only identified prior research that included a cueing procedure for the Visual Reproduction subtest was conducted by Gass (1995). In that study, the examiner provided a visual cue by slowing drawing a small segment of each design. The examiner continued to draw the design until the participant indicated that they were able to complete it. However, this procedure could potentially result in each participant receiving different "cues". Later recognition testing may also be confounded depending on the extent of the "cue" drawn by the examiner, which may be different across participants.

In this study an alternative standardised procedure for cueing was developed. Rather than continuing to draw a small segment of the design, all participants were presented with the same partial cue that did not represent the whole design. This was considered to be a more reliable form of cueing because each participant received the same partial cue, the partial cue had less potential to confound later recognition performance and presentation of the same partial cue did not differentially benefit some participants.

In contrast to the Gass (1995) procedure, any recall with the assistance of a cue was scored without prejudice for the cue in order to reflect memory recall. Gass only scored those portions of the design that the participant produced. However, to only score the participants recall could be seen as implying that they had encoded/recalled only those portions of the design rather than the visual image as an entirety. It could be argued that if a person's memory is prompted by a cue, all the material should be scored, not just those that the participant recalled. If the information of interest is the retention of information, not just retrieval, then scoring all elements is likely to provide a better indication of what was retained.

More than 50% of the sample in the current study required a cue on more than one design because they failed to recall on delayed free recall. This included both control and experimental participants, although more experimental participants required a cue (80% vs 47%). Additionally, experimental participants generally required cueing on more than one design as compared to the control participants. When cueing was provided, more than half of the participants were able to recall information about the design (66% of the experimental participants and 86% of the control participants). Indeed, when further recall was produced, there would appear to have been a significant score advantage obtained for the group as a whole. These results suggested that the provision of cues was able to facilitate further memory recall in many cases when the participant had no previous recall. This data supports the idea that individuals acquire more information than may be disclosed by measures of free recall.

Cued recall performance in this study suggests that those with and without clinical pathology encode and store more information than what may be produced on free recall. Members of the control group benefited from a cue in this study not just those with cerebral lesions. Thus, benefiting from a cue may be a normal phenomenon and not just one seen in a clinical population. If this is the case, interpreting the result of cueing as reflecting a deficit in encoding or storage may be a premature. An alternative explanation is that those in the control group who required cueing represented a sub-clinical population, who had very mild memory impairment. Hence, the diagnostic significance of the benefit from a cue needs to be explored further.

The results found in the current study were similar to those found by Gass (1995). He found that fifty-five percent of his sample required cueing to recall at least one design. Those who required cueing scored an average of five additional points (or 34% more detail). Twenty percent of participants in Gass's study were reported to have increased their recall score by over 50%. Cued recall also provided better discrimination between a psychiatric and neurologically compromised persons than did free recall. Thus, the findings of the current study and Gass's study suggest that participants acquire more information than may be implied by performance on free recall. To interpret memory performance based solely on free recall is therefore likely to be somewhat misleading.

The cueing system developed in this study has a number of advantages. Firstly, only a partial cue is provided and therefore it is easier to identify what people were able to recall. Secondly, there is no interaction between the examiner and participant when providing the cue. Thus, the examiner cannot inadvertently further cue the participant by other means, making the process more reliable. Thirdly, a partial cue is only given for the designs not recalled and therefore the procedure is economical in terms of time. Fourthly, the use of a cue examines the minimum assistance required to prompt recall rather than going straight to recognition. Finally, the presentation of the cue does not confound subsequent recognition performance. This is important because following cueing with recognition enriches memory testing for those participants who have no delayed recall and do not benefit from cueing. Poor performance on cueing does not imply in itself an impairment in retrieval.

Although further study is needed to explore the diagnostic implications of cued recall on the Visual Reproduction subtest, what is evident from this study is that a substantial number of individuals may be able to recall more information with the benefit of a cue. This has a number of potential applications, particularly in the rehabilitation setting.

Recognition- Contribution of Encoding and Retrieval to Memory Recall

Recognition typically involves the participant selecting or identifying an item as being one that they learned previously. If performance is superior on a recognition task, as compared to free recall, a retrieval deficit may be present. If performance is equally poor on both recall and recognition, then an encoding problem or storage deficit is more likely, depending on whether the task is one of immediate recall or delayed recall.

The WMS-R version of the Visual Reproduction subtest had no standardised procedure for assessing recognition memory. The WMS-111 version of the Visual Reproduction subtest did incorporate a recognition task, however, there were a number of potential problems with the format of the task, notably that it was lengthy and time consuming, that the designs in Items 4 and 5 were presented separately which could confuse participant, that the distracters were similar to the targets which could result in incorrect partial recognition and that guessing could confound interpretation of performance. The Boston Revision of the WMS (Milberg et al., 1986) also incorporated immediate recognition and delayed recognition trials. However, the additional exposure to the designs from the immediate recognition task and the copy trials may confound interpretation of delayed recall because of the extra exposure to the designs. Thus, previous attempts at incorporating a recognition task into the standard Visual Reproduction procedure have been problematic.

In this study, recognition was assessed with a six alternative-choice recognition task. The participant was required to identify the correct stimuli from five distracters. This task was similar to that reported by Fastenau (1996) although the actual stimuli comprising his recognition task were not presented in his study. That task incorporated five choices and Fastenau reported limited reliability and weak psychometric properties.

The results on the delayed recognition task administered in the current study showed that the majority of the control group had perfect recognition. This was also true for the clinical group, although 33% of the clinical group did have less than perfect recognition. A majority of the clinical participants who had no recall of one or more designs on delayed or cued recall were able to recall additional information with the assistance of the recognition format. This indicates that a number of the participants in this study were able to recognise material that they were not able to recall freely on delayed recall or with the assistance of a cue. Thus, administration of a recognition trial in this study generated additional information of potential clinical significance.

Although many participants were able to recall more information with the assistance of a recognition task, there was a small number of the total sample that did not (13%). However, there were only three participants in the clinical group who were unable to recognise any of the designs that they had not previous recalled. Many of the experimental participants who did not recognise a design had no free recall and were not assisted by a cue. Hence, impairments in encoding and storage were suggested, not just limitations in retrieval.

The psychometric properties of the new procedure were generally sound. Results suggested that the recognition procedure had marginally better internal consistency than that of Fastenau (1996). Although the internal consistency was still not high, the small number of items was likely to have influenced the strength of the reliability coefficient. The internal consistency was also likely to have been affected by the ceiling effects in both the clinical and experimental groups.

In addition to affecting internal consistency, the ceiling effects evident on the multiple choice task are likely to result in difficulty making discriminations between high scores on this task. However, because most participants score perfectly on recognition this is unlikely to be a large problem. Instead, low scores could be useful in identifying people with memory difficulties.

In this study, there was also a strong relationship between performance on the recognition task and free recall. There was a moderate relationship between free recall and recognition in the control group (Immediate rho=.33; Delayed rho=.40) and a large relationship in the experimental group (Immediate rho=.59; Delayed rho=.66). Thus, a high score on free recall is likely to be associated with a high score on the recognition task. This provides support for the concurrent validity of the recognition measure.

Scores on the recognition task were significantly different between the control and the experimental groups, with the control group performing better on this task. This is not suprising since the control group had better free recall and were therefore more likely to be able to recognise the designs in a multiple-choice format. Further work will need to be conducted to identify if there are certain subgroups of people who are more likely to benefit from this recognition procedure on the Visual Reproduction subtest. As research has shown that there are some conditions which are characterised by good recognition performance and much weaker free recall, it would be useful to establish the validity and utility of the recognition procedure in a variety of groups.

Certainly this study revealed that more information can be obtained about memory performance when using a recognition format. The actual procedure also has a number of advantages. Firstly, it requires very little administrative time as the recognition procedure need only be applied to those designs where there is no recall or minimal recall. Secondly, because there are six items in the recognition task, the procedure has a better ability to discriminate benefit from recognition as compared to chance performance. Thirdly, administering the task does not alter the standard administration, which means that the original norms can be used when evaluating immediate and delayed recall. Fourthly, because the correct design and the distracters are all presented at the same time there is no demand on memory during the evaluation of this function. Thus, the participant does not have to remember a number of distracters in order to make a decision about which one was the correct stimuli. This also means that the task is not going to be contaminated by material that was previously presented, because all the material necessary to make the recognition judgement is present in front of the participant.

Taken together, the findings from the cueing procedure and multiple choice tasks suggest that many participants acquire substantially more information than that which may be suggested by measures of free recall. Thus, the use of these tasks can generate more clinical information than the standard administration and the findings suggest that it is important that all participants are given a variety of opportunities to demonstrate what they have stored in memory. This opportunity would involve the routine administration of additional memory procedures such as the ones designed in this study. However, as Fastenau (1996) suggested, there is a need for a wide-ranging study across different ages and different clinical groups to develop useful normative data.

Perceptual Match- Contribution of Perception to Memory Recall

One of the major problems with the Visual Reproduction subtest is that impairments in construction, perception or motor abilities can potentially influence the scoring and the interpretation of performance on the task (Gfeller et al., 1995; Haut et al., 1994; Ricker et al, 1994). The task has been shown to share strong associations with visuo-constructional and visuo-perceptual tasks (Trahan et al., 1988). Indeed, the current study the Visual Reproduction subtest shared a strong correlation with the Block Design subtest of the WAIS-R. However, this does not necessarily mean that Visual Reproduction is simply a measure of non-verbal cognitive ability. It does suggest, however, that there is a large non-verbal cognitive component, and this should be kept in mind when interpreting performance on the task. Fortunately, procedures can be developed in order to understand the contribution of visuo-construction and visual perception to performance on memory tasks.

Impaired visual-perceptual skills may result in poor performance on the Visual Reproduction task because it could compromise the accurate encoding of visual information. In this study, there were some participants who were unable to recall material despite being provided with a cue and a recognition procedure. Since perceptual processes are involved in the task, it was important clinically to establish that poor recall and recognition was not being confounded by perceptual problems. Indeed, it has been found that performance on the Visual Reproduction subtest can be potentially influenced by visual perceptual difficulties (Ricker et al., 1994). A task that involves matching the correct visual reproduction design would be of use in identifying participants whose impaired visual perceptual skills contributed to their defective performance on a visual memory task.

The perceptual match task in this study was similar to that used by Fastenau (1996) and the procedure included in the WMS-111. Each design was presented with six similar designs and the participant had to match the same designs. The results of this study suggested that poor matching was rare, even in a clinical group. However, some participants did do poorly on this task. Those participants who matched less than three of the designs correctly had minimal of no recall of the designs with any other procedure. Poor matching following deficient design recall is likely to implicate perceptual or attentional difficulties rather than a primary memory disorder. This suggests that for these participants visual-perceptual difficulties could have affected their memory recall. In these cases, it may have been wise to exclude their results from the other analysis, because visual perceptual skills may have contributed to memory performance. This may reduce the relationship between the Visual Reproduction task and other memory memories, in addition to confounding group differences. Further research will need to consider excluding data for those participants who do poorly on perceptual in order to better understand the nature of the Visual Reproduction subtest as a memory measure.

The perceptual match subtest in this study had poor internal consistency. Similarly, Fastenau (1996) found that his perceptual match task had limited reliability. However, the reduced number of items and the reduced range of scores in the sample used are likely to have impacted on the strength of the reliability quotients derived. However, the perfect matching scores obtained by nearly all the control group suggest that a low score on the task is potentially very meaningful. Further research should focus on developing normative data for a number of clinical conditions in order to provide clinicians with a basis for understanding the relationship between matching and recall.

Future Directions

Although the revised scoring system is a potentially useful addition to the standard Visual Reproduction protocol it will be necessary to carry out a more extensive analysis of its psychometric qualities and to build up normative data. Information is required about the ability of the revised scoring system to make finer distinctions about recall quality and about the reliability in scoring with the revised system in different populations and by different scorers.

Although the two scoring systems produce a similar grading of memory performance, the results of this study cannot be used to argue that the revised scoring system provides a better grading of memory recall. Further research would need to be conducted in this regard. This could potentially involve experienced clinicians, niave to the purpose of the study, grading the quality of memory recall based on clinical judgement and comparing this to scores obtained on the original scoring system and the revised scoring system.

The non-verbal memory Index developed in this study has the potential to be able to identify participants with primarily non-verbal memory impairment. However, this Index was based on a small sample. Thus, further research will be required to validate the Index in larger samples. Data will also need to be collected to identify whether the new Index is able to quantify the severity of non-verbal memory impairment.

The development of the additional measures in this study supported the idea that more information about memory functioning can be extracted by supplementing standard tests with additional procedures. Further research will need to be conducted to provide a normative base and to identify if different patterns of performance on these tasks have diagnostic value. These studies could look more broadly at the use of cueing and recognition in non-clinical and clinical populations to provide a better understanding of the significance of the improvements individuals make.

Although this study included cueing, recognition and perceptual match tasks, these do not provide direct information about contribution of drawing and construction to design recall. The revised scoring system was designed to reduce the impact of drawing difficulties on memory performance, however constructional difficulties are less controlled. Although a recognition task could be of potential use for participants with impaired constructional skills, in the format given in this study it would be difficult to ascertain if better recognition was the result of poor constructional ability or poor free recall. A copy trial is likely to be most beneficial to understanding recall of designs.

The WMS-111 version of the WMS-111 included a copy trial, however it did not provide a quantitative method to compare it with the recall trials in order separate memory of material from constructional dysfunction. Similarly, although Fastenau (1996) also used a copy trial in his study, he did not present a standardised way of understanding the contribution of defective visuo-construction skills to memory performance. Haut et al. (1996) developed a method for quantifying the contribution of visual construction to Visual Reproduction performance by using a copy trial. Further development of the Visual Reproduction test as a comprehensive memory tool will need to include a procedure to evaluate the copy of the designs in comparison to memory recall. A copy trial would not be inconsistent with the standardised administration of the Visual Reproduction subtest and the administration of additional measures.

Conclusion

The primary goal of this study was to develop a revised scoring system for the Visual Reproduction of the WMS-R. Results indicated that when the revised scoring system and the original scoring system were applied rigorously, they were very highly correlated in an elderly population and therefore generated a similar grading of memory. The results also indicated that the revised scoring system had sound psychometric properties, including very high inter-rater and intra-rater reliability and strong internal consistency. The revised scoring system had at least equivalent reliability to the original scoring system and was also found to generate a wider range of scores. Although the two scoring systems produce a similar grading of memory performance, the results of this study cannot be used to argue that the revised scoring system provides a better grading of memory recall.

Due to the very high association between the two scoring systems, the revised system could replace the original system without any substantial change in the grading of memory performance. In addition, the revised scoring system developed was found to be at least as reliable as the original and did not appear to introduce new problems that might compromise its' clinical utility in measuring memory function. Because it does not require any changes to the standard administration, the transition to using the revised scoring system less problematic. Thus, the high association between the two scoring systems and the good psychometric properties supported the approach to developing a revised scoring system, that is, to eliminate/reduce anomalies with the original and reduce the emphasis on the aspects of performance that may not be reflective of memory function. Further research is needed to extend the psychometric and clinical data obtained in this study.

This study also evaluated the revised scoring system as a measure of non-verbal memory. Data from the control group supported the distinction between tasks measuring non-verbal and verbal memory, although all memory measures shared some relationship. The basis of the Visual Reproduction subtest was further examined by comparing the difference between two groups, one with left hemisphere damage and one with right hemisphere damage. Results were consistent with the material specific model of memory, with lateralised right hemisphere damage being associated with significantly poorer performance on the Visual Reproduction subtest, when scored by both the original and revised scoring systems.

A non-verbal memory Index was developed based on items on the revised scoring system that differentiated between groups with lateralised cerebral damage. This Index was shown to have good sensitivity when identifying those with right hemisphere lesions. However, since the Index was derived from the same data as it was validated against (the left and right group), the discriminating power of the Index may have been overestimated. This study also did not provide information about the severity of the non-verbal deficit based on the Index. Further work will need to establish if this Index has sensitivity and specificity in identifying people with non-verbal memory impairment in other sample. The use of an Index rather than a total score to identify memory deficits may be a more fruitful strategy to investigating non-verbal memory. This approach recognises that both verbal and non-verbal encoding of visual material takes place.

Cueing, recognition and perceptual match procedures were also developed and standardised in this study with the goal of providing more information about memory from this task. Although further study is needed to examine whether performance on cued recall and recognition has diagnostic implications in some cases, what is evident from this study is that a number of people are able to recall more information than is implied by free recall. Thus, the use of these tasks generates more information than the standard administration. As a result, administration of the Visual Reproduction subtest should involve the routine administration of cueing and recognition trials so that all participants are given adequate opportunity to demonstrate their best memory recall. However, normative data is required before the procedures can have wide application.

REFERENCES

Adams, R. (1989). Review of the recognition memory test. In J. L. Conoley & J. J. Kremer (Eds.), *The 10th Mental Measurements Yearbook* (pp. 693-694). Lincoln: Buros Institute of Mental Measurement.

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.

Atkinson, R. C. & Shiffrin, R. M. (1968). 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. 82-90). New York: Academic Press.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D. (1990). *Human memory.* Hove, England: Erlbaum.

Baddeley, A. D. (1995) The psychology of memory. In A. D. Baddeley, B. A. Wilson & F. N. Watts (Eds.), *Handbook of memory disorders (*pp. 3-26). Chichester: John Wiley and Sons.

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. 47-90). San Diego: Academic Press.

Barr, W. B., Chelune, C. J., Hermann, B. P., Loring, D. W., Perrine, K., Strauss, E., Trennery, 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.

Baxendale, S. A. (1997). The role of the hippocampus in recognition memory. *Neuropsychologia, 35,* 591-598.

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

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

Bigler, E. D., Johnson, S. C., Anderson, C. V., Blatter, D. D., Gale, S. D, Russo, A. A., Ryser, D. K., Macnamara, 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 (London), 232,* 331-356.

Boller, F. & De Renzi, E. (1967). Relationship between visual memory defects and hemispheric locus of lesion. *Neurology, 17,* 1052-1058.

Bondi, M. W., Salmon, D. P. & Kaszniak, A. W. (1996). The neuropsychology of dementia. In I. Grant and K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp. 164-199). New York: Oxford University Press.

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

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., Pakalnis, A., & Drake, M. E., Jr. (1988). Verbal and nonverbal memory and learning in patients with complex partial seizures of temporal lobe origin. *Journal of Epilepsy, 1,* 203-208.

Bowden, S. C. (1997). Factor structure of the Wechsler Memory Scale-Revised in people with hippocampal sclerosis. *Australian Journal of Psychology, 49 (Suppl.),* 134.

Bowden, S. C., Dodds, B., Whelan, G., Long, C., Dudgeon, P., Ritter, A. & Clifford, C. (1997). Confirmatory factor analysis of combined Wechsler Memory Scale-Revised in a sample of clients with alcohol dependency. *Psychological Assessment, 11,* 339-344.

Bowden, S. C., Carstairs, J. R., & Shores, E. A. (1999). Confirmatory factor analysis of combined Wechsler Adult Intelligence Scale- revised and Wechsler Memory Scale-Revised scores in a healthy community sample. *Psychological Assessment, 11,* 339-344.

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The Clinical Neuropsychologist, 5,* 125-142.

Brandt, J. & Rich, J. B. (1995). Memory disorders in the dementias. In A. D. Baddeley, B. A. Wilson & F. N. Watts (Eds.), *Handbook of memory disorders* (pp. 243-270). Chichester: John Wiley and Sons.

Burton, D. B., Mittenberg, W. & Burton, C. A. (1993). Confirmatory factor analysis of the Wechsler Memory Scale- Revised standardisation sample. *Archives of Clinical Neuropsychology*, *8,* 467-475.

Buschke, H., & Fuld, P. A. (1974). Evaluation of storage, retention and retrieval in disordered memory and learning. *Neurology, 11,* 1019-1025.

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., Salmon, D. P., Cullum, C. M., Carins, P. Troster, A. I., Jacobs, D., Moss, M., & Cermack, L. W. (1988). Differentiation of amnesic and demented patients with the Wechsler Memory Scale- Revised. *The Clinical Neuropsychologist, 2,* 133-148.

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

Carlesimo, G. A., & Caltagirone, G. (1995). Components in the visual processing of known and unknown faces. *Journal of Clinical and Experimental Neuropsychology, 17,* 691-705.

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

Cermack, L. S., & Tarlow, S. (1978). Aphasic and amnesic patients' verbal vs. nonverbal retentive abilities. *Cortex, 14,* 32-40.

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

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,* 399-404.

Cohen, N. J. & Eichenbaum, H. (1993). *Memory, amnesia and the hippocampal system.* London, England: MIT Press, A Bradford Book.

Cohen, N. J. & Squire, L. R. (1980). Preserved learning and retention of pattern analysing skill in amnesia: Dissociation of knowing how and knowing what. *Science, 210,* 207-209.

Corkin, S. (1968). Lasting consequences of bilateral medial temporal lobectomy: clinical course and experimental findings in H.M. *Seminars in Neurology, 4,* 249-259.

Corkin, S. (1984). Acquisition of motor skills after bilateral medial temporal lobe excision. *Neuropsychologia, 6,* 225-265.

Craik, F., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal learning and Verbal Behaviour, 11,* 671-684.

Cummings, J. L. (Ed.) (1990). *Subcortical dementia.* New York: Oxford University Press.

Curran, T. & Schacter, D. L. (1997). Amnesia: Cognitive neuropsychological aspects. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioural neuropsychology and neuropsychology* (pp. 463-471). New York: McGraw Hill.

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.

Delaney, R. C., Rosen, A., Mattson, R. H., & Novelly, R. A. (1980). Memory function in focal epilepsy: A comparison of non-surgical, unilateral temporal lobe and frontal lobe samples. *Cortex, 16,* 103-117.

D'Elia, L. D., Satz, P., & Schretlen, D. (1989). Wechsler Memory Scale: A critical appraisal of the normative studies. *Journal of Clinical and Experimental Neuropsychology, 11,* 551-568.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult Version.* San Antonio, TX: The Psychological Corporation.

Denman, S. (1984). *Denman Neuropsychology Memory Scale.* Charleston, SC: S. B. Denman.

Des Roziers, G., & Ivison, D. (1986). Paired associate learning: Normative data for differences between high and low associate word pairs. *Journal of Clinical and Experimental Neuropsychology, 8,* 637-642.

Dobbins, I. G., Kroll, N. F. A., Tulving, E., Knight, R. T., & Gazzzniga, M. S. (1998). Unilateral medial temporal lobe memory impairment: type deficit, function deficit or both? *Neuropsychologia, 36,* 115-127.

Eadie, K. & Shum, D. (1995). Assessment of visual memory: A comparison of Chinese characters and geometric figures as stimulus materials. *Journal of Clinical and Experimental Neuropsychology, 17,* 731-739.

Elwood, R. W. (1991a). The Wechsler Memory Scale-Revised: Psychometric characteristics and clinical application. *Neuropsychology Review, 2,* 179-201.

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

Elwood, R. W. (1993). Delayed recall on the Wechsler Memory Scale-Revised: The factor structure revisited. *Journal of Clinical Psychology, 49,* 854-859.

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

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

Feher, E. P. & Martin, R. C. (1992). Cognitive assessment of long-term memory disorders. In D. I. Margolin (Ed.), *Cognitive neuropsychology in clinical practice* (pp. 168-203). New York: Oxford University Press.

Fischer, J. S. (1988). Using the WMS-R to detect and characterise deficits in multiple sclerosis. *The Clinical Neuropsychologist, 2,* 149-172.

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

Gabrieli, J. D. E., Cohen, N. J., & Corkin, S. (1988). Impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and Cognition, 7,* 157-177.

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

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

Giovagnoli, A. R. & Avanzini, G. (1999). Learning and memory impairment in patients with temporal lobe epilepsy: Relation to the presence, type and location of brain lesion. *Epilepsia, 40,* 904-911.

Gliebner, U., Helmstaedter, C., & Elger, C. E. (1998). Right hippocampal contribution to visual memory: a presurgical and postsurgical study in patients with temporal lobe epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry, 65,* 665-669.

Glosser, G, Goodglass, H, & Biber, C. (1989). Assessing visual memory disorders. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 82-91.

Glosser, G, Deutsch, G. K., Cole, D. C., Corwin, J., & Saykin, A. J. (1998). Differential lateralisation of memory discrimination and response bias in temporal lobe epilepsy patients. *Journal of the International Neuropsychological Society, 4,* 502-511.

Graf, P. & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 501-518.

Hanger, P. A., Montague, J. R., & Smith, M. (1991, Feb). A recognition memory test for the visual reproduction subtest of the WMS-R. *Paper presented at the meeting of the International Neuropsychology Society*, Galveston, Tx.

Hannay, H. J., Levin, H. S., & Grossman, R. J. (1979). Impaired recognition memory after head injury. *Cortex, 15,* 269-283.

Haut, M. C., 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,* 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. *The Clinical Neuropsychologist, 8,* 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,* 299-309.

Helmstaedter, C., & Elger, C. E. (1996). Cognitive consequences of two-thirds anterior temporal lobectomy on verbal memory in 144 patients: A three month follow-up study. *Epilepsia, 37,* 171-180.

Helmstaedter, C., Hufnagel, A., & Elger, C. E. (1992). Pre-operative memory profiles in patients with temporal lobe epilepsy are related to post-operative seizure outcome. *Journal of Epilepsy, 5,* 17-32.

Helmstaedter, C., Pohl, C., & Elger, C. E. (1995). Relations between verbal and nonverbal memory performance: Evidence of confounding effects particularly in patients with right temporal lobe epilepsy. *Cortex, 31,* 345-355.

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, D. O. (1988). Development of the Wechsler Memory Scale- Revised. *The Clinical Neuropsychologist, 2,* 102-106.

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

Herman, B. P., Seidenberg, M., Haltiner, A., & Wyler, A. R. (1995b). Relationship of age at onset, chronological age and adequacy of preoperative performance to verbal memory change after anterior temporal lobectomy. *Epilepsia, 36,* 137-145.

Herman, B. P., Wyler, A. R., & Somes, G. (1993). Memory loss following left anterior lobectomy is associated with hippocampal pathology and not the extent of hippocampal resection. *Journal of Clinical and Experimental Neuropsychology, 15,* 24.

Herman, B. P., Wyler, A. R., Somes, G., Berry 111, A.D., & Dohen, F. C., Jr. (1992). Pathological status of the mesial temporal lobe predicts memory outcome from left anterior temporal lobectomy. *Neurosurgery, 31,* 652-657.

Hodges, J. R. (1995). Retrograde amnesia. In A. D. Baddeley, B. A. Wilson & F. N. Watts (Eds.), *Handbook of memory disorders (*pp. 81-108). Chichester: John Wiley and Sons.

Hodges, J. R., Salmon, D. P. & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia, 30,* 301-314.

Hunkin, N. M., Stone, J. V., Isaac, C. L., Holdstock, J. S., Butterfield, R., Wallis, L. I., & Mayes, A. R. (2000). Factor analysis of three standardised tests of memory in a clinical population. *British Journal of Clinical Psychology, 39,* 169-180.

Ivison, D. (1993). Towards a standardisation of the Wechsler Memory Scale Form 2. *The Clinical Neuropsychologist, 3,* 268-280.

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

Jones, R. D. & Tranel, D. (1993). Autobiographical memory in patients with focal brain lesions. *The Clinical Neuropsychologist, 7,* 340.

Jones-Gottman, M. (1986a). Memory for designs: The hippocampal contribution. *Neuropsychologia, 24,* 193-203.

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

Jones- Gottman, M. (1987). Commentary: Psychological evaluation- testing hippocampal function. In J. Engel (ed.), *Surgical treatment of the epilepsies* (pp. 511-540). New York: Raven Press.

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

Kandel, E. R., & Schwartz, J. H. (1982). Molecular biology of learning: modulation of transmitter release. *Science, 218,* 433-443.

Kapur, N. (1987). Some comments on the technical acceptability of Warrington Recognition Memory Test. *The British Journal of Clinical Psychology, 26,* 144-146.

Kapur, N. (1988). *Memory disorders in clinical practice.* London: Butterworths.

Kaufman, A. S. (1990). *Assessing Adolescent and Adult Intelligence.* Boston: Allyn and Bacon.

Kimura, D. (1963). Right temporal lobe damage. *Archives of Neurology, 8,* 48-55.

Kopelman, M. D. (1992). The psychopharmacology of human memory disorders. In B. Wilson & N. Moffat (Eds.), *Clinical Management of memory problems* (2nd Ed., pp. 189-215). London: Chapman and Hall.

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

Kopelman, M. D. (1995b). The assessment of psychogenic amnesia. In A. D. Baddeley, B. A. Wilson & F. N. Watts (Eds.), *Handbook of memory disorders (*pp. 427-450). Chichester: John Wiley and Sons.

Kopelman, M. D. (1998). Clinical and neuropsychological studies of patients with amnesic disorders. In M. A. Ron & A. S. David (Eds.), *Disorders of brain and mind* (pp. 147-174). Cambridge: Cambridge University Press.

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

Larrabee, G. J., Kane, R. L. & Schuck, J. R. (1983). Factor analysis of the WAIS and Wechsler Memory Scale: An analysis of 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.

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

Lencz, T., McCarthy, G., Bronen, R. A., Scott, T. M., Inserni, J. A., Sass, K. J., Novelly, R. A., Kim, J. H., & Spencer, D. D. (1992). Quantitative magnetic reasoning imaging in temporal lobe epilepsy: relationship to neuropathological and neuropsychological function. *Annals of Neurology, 31,* 629-637.

Leng, N. R. C. & Parkin, A. J. (1990). The assessment of memory disorders: A review of some current clinical tests. *Clinical Rehabilitation, 4,* 159-165.

Leonberger, F. T., Nicks, S. D., Goldfader, P. R., & Munz, D. C. (1991). Factor analysis of the Wechsler Memory Scale-Revised and the Halstead-Reitan Neuropsychological Battery. *Clinical Neuropsychologist, 5,* 83-88.

Leonberger, F. T., Nicks, S. D., Larrabee, G. J., & Goldfader, P. R. (1992). Factor structure of the Wechsler Memory Scale-Revised within a comprehensive neuropsychological battery. *Neuropsychology, 6,* 239-249.

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

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

Loring, D. W., Lee, G. P., Martin, R. C., & Meador, K. J. (1988). Material specific learning in patients with complex seizures of temporal lobe origin: Convergent validation of memory constructs. *Journal of Epilepsy, 1,* 53-59.

Loring, D. W., Lee, G. P., Martin, R. C., & Meador, K. J. (1989b). Verbal and Visual Memory Index discrepancies from the Wechsler Memory Scale-Revised: Cautions in interpretation. *Psychological Assessment, 1*, 198-202.

Loring, D. W., Lee, G. P.,& Meador, K. J. (1989c). Issues in memory assessment of the elderly. *Clinics in Geriatric Medicine, 5,* 565-581.

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.

MacCartney-Filgate, M. S., & Vriezen, E. R. (1988). Intercorrelation of clinical tests of verbal memory. *Archives of Clinical Neuropsychology, 3,* 121-126.

Malec, J. F., Ivnik, R. J., & Hinkeldey, N. S. (1991). Visual spatial learning test. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 82-88.

Mandler, G. (1980). Recognising: The judgement of previous occurrence. *Psychological Review, 87,* 252- 271.

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). Chichester: John Wiley and Sons.

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.

McCarthy, R. A. & Warrington, E. K. (1990). *Cognitive neuropsychology: A clinical introduction.* San Diego: Academic Press, Inc.

McGeer, P. L., Eccles, J. C. & McGeer, E. D. (1978). *Molecular neurobiology of the mammalian brain.* New York: Plenum Press.

McGuire, B. E. & Batchelor, J. (1998). Inter-rater reliability of the WMS-R logical memory and visual reproduction subtests in a neurosurgical sample. *Australian Psychologist, 33,* 231-233.

McMillan, T. M., Powell, G. E., Janota, I. & Polkey, C. E. (1987). Relationships between neuropathology and cognitive functioning in temporal patients. *Journal of Neurology, Neurosurgery and Psychiatry, 50,* 167-176.

McPherson, S. & Cummings, J. L. (1996). Neuropsychological aspects of Parkinson's Disease and Parkinsonism. In I. Grant and K. M. Adams (Eds.), *Neuropsychological Assessment of Neuropsychiatric Disorders* (2^{nd} Edn.) (pp. 288-310). New York: Oxford University Press.

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 and K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp. 56-86). New York: Oxford University Press.

Millis, S. R., Malina, A. C., Bowers, D. A., & Richer, J. H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale-111. *Journal of Clinical and Experimental Neuropsychology, 21,* 87-93.

Milner, B. (1958). Psychological deficits produced by temporal lobe excision. *Research Publications for the Association of Research in Nervous and Mental Disease, 36,* 244-257.

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, (1968). Visual recognition and recall after right temporal lobe excision in man. *Neuropsychologia, 6,* 191-209.

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

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

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London.* Series B: Biological Sciences, 298, 85-92.

Mitrushina, M. N., Boone, K. B., & D'Elia, L. F. (1999). *Handbook of normative data for neuropsychological assessment* (pp. 267-322). New York: Oxford University Press.

Morris, R. G., Abrahams, S., Baddeley, A. D. & Polkey, C. E. (1995a). Doors and People: Visual and verbal memory after unilateral temporal lobectomy, *Neuropsychology, 9,* 464-469.

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

Murdock, B. B. (1982). Recognition memory. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 1-25). New York: Academic Press.

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

Naugle, R. I., Chelune, G. J., Luders, H., & Awad, I. A. (1991). Detection of changes for material-specific information following left versus right temporal lobe lobectomy using the Wechsler Memory Scale-Revised. *Journal of Clinical and Experimental Neuropsychology, 13,* 45. (Abstract)

Naugle, R. I., Chelune, G. J., Schuster, J., Luders, H., & Comair, Y. (1994). Recognition memory for words and faces before and after temporal lobectomy. *Assessment, 1,* 373-381.

Newcombe, E., de Haan, E. H. E., Ross, J., & Young, A. W. (1989). Face processing, laterality and contrast sensitivity. *Neuropsychologia, 27,* 523-538.

Nicks, S. D., Leonberger, F. T., Munz, D. C., & Goldfader, P. R. (1992). Factor analysis of the WMS-R and the WAIS. *Archives of Clinical Neuropsychology, 7,* 387-393.

O'Connor, M., Verfaellie, M. & Cermack, L. S. (1995). Clinical differentiation of amnesic subtypes. In A. D. Baddeley, B. A. Wilson & F. N. Watts (Eds.), *Handbook of memory disorders (*pp. 53-80). Chichester: John Wiley and Sons.

Oresick, R. J. & Broder, S. N. (1988). The psychometric structure of the WMS in comparison to the WAIS-R in a low-IQ clinical population. *Journal of Clinical Psychology, 44,* 549-557.

Parkin, A. (1984). Amnesic syndrome: A lesion specific disorder? *Cortex, 20,* 479-508.

Parkin, A. J. & Leng, N. R. (1993). *Neuropsychology of the Amnesic Syndrome* (pp. 17-34). Hove: Lawrence Erlbaum Associates.

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.

Powell, G. E., Polkey, C. E., & McMillan, T. (1985). The new Maudsley series of temporal lobectomy 1: Short term cognitive effects. *British Journal of Clinical Psychology, 24,* 109-124.

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

Rapaport, D., Gill, M. M., & Schafer, R. (1968*). Diagnostic Psychological Testing* (Rev. Edn.), R. R. Holt (Ed.) New York: International University Press.

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.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie, 28,* 286-340.

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

Ricker, J. H., Keenman, P. A., & Jacobsen, M. W. (1994). Visuoperceptual-spatial ability and visual memory in vascular dementia and dementia of the Alzheimer's type. *Neuropsychology, 32,* 1287-1296.

Roid, G. H., Prifitera, A. & Ledbetter, M. (1988). Confirmatory analysis of the factor structure of the Wechsler Memory Scale- Revised. *The Clinical Neuropsychologist, 2,* 116-120.

Roth, D. L., Conboy, T. J., Reeder, K. P., & Boll, T. J. (1990). Confirmatory factor analysis of the Wechsler Memory Scale-Revised in a sample of head-injured patients. *Journal of Clinical and Experimental Neuropsychology, 12,* 834-842.

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

Russell, E. W. (1988). Renorming Russell's version of the Wechsler Memory Scale. *Journal of Clinical and Experimental Neuropsychology, 10,* 235-249.

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

Saling, M. M. (1998). *The road less travelled.* A Workshop presented at the Fourth National Conference of the APS College of Clinical Neuropsychologists. Lorne, Victoria, Australia.

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.

Saling, M. M., O'Shea, M., & Berkovic, S. F. (1995). Verbal memory in temporal lobe epilepsy: Cognitive influences and task-specific effects. *Epilepsia, 36 (Suppl 3),* 93-94.

Sass, K. J., Sass, A., Westerveld, M., Lencz, T., Rosewater, K. M., Novelly, R. A., Kim, J. H. & Spencer, D. D. (1992). Russell's adaptation of the Wechsler Memory Scale as an index of hippocampal pathology. *Journal of Epilepsy, 5,* 24-30.

Saykin, A. J., Gur, R. C., Sussman, N. M., O'Connor, M. J. & Gur, R. E. (1989). Memory deficits before and after temporal lobectomy: Effects of laterality and age of onset. *Brain and Cognition, 9,* 191-200.

Senior, G. and Douglas, L. (1999). *Issues and methods in medicolegal assessment.* A Workshop presented at the Fifth National Conference of the APS College of Clinical Neuropsychologists. Leura, NSW, Australia, October, 1999.

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

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22,* 261-273.

Shepherd, G. M. (1994). *Neurobiology*, 3rd Edn. NewYork: Oxford University Press.

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

Skilbeck, C. E. & Woods, R. T. (1980). The factorial structure of the Wechsler Memory Scale: Samples of neurological and psychogeriatric patients. *Journal of Clinical Neuropsychology, 2,* 293-300.

Smith, G. E., Ivnik, R. J., Malec, J. F., Korkmen, E., Tangalos, E. G. & Kurland, L. T. (1992a). Mayo's older American normative studies (MOANS): Factor structure of a core battery. *Psychological Assessment, 4,* 382-390.

Smith, G. E., Ivnik, R. J., Malec, J. F., Peterson, R. C. Korkmen, E. & Tangalos, E. G. (1994). Mayo cognitive factor scales: Derivation of a short battery & norms for factor scores, *Neuropsychology, 8,* 194-202.

Smith, G. E., Ivnik, R. J., Malec, J. F., & Tangalos, E. G. (1993). Factor structure of the Mayo older americans normative sample (MOANS) core battery: Replication in a clinical sample. *Psychological Assessment, 5,* 121-124.

Smith, M. L., Malec, J. F., & Ivnik, R. J. (1992b). 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.

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

Snowden, J. S., Goulding, P. J. & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural Neurology, 2,* 167-182.

Snowden, J. S., Griffiths, J. & Neary, D. (1994). Semantic dementia: Autobiographical contribution to preservation of meaning. *Cognitive Neuropsychology, 11,* 265-288.

Snowden, J. S., Neary, D. & Mann, D. M.A. (1996). *Fronto-temporal lobar degeneration: Fronto-temporal dementia, progressive aphasia, semantic dementia.* New York: Churchill and Livingston.

Snowden, J. S., Neary, D., Mann, D. M. A., Goulding, P. J. & Tsta, H. J. (1992). Progressive language disorder due to lobar atrophy. *Annals of Neurology, 31,* 174-183.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms and commentary.* New York: Oxford University Press.

Squire, L. R. (1982). The neuropsychology of human memory. *Annual Review of Neuroscience, 5,* 241-273.

Squire, L. R. (1986). Mechanisms of memory, *Science, 232,* 1612-1619.

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

Squire, L. R. & Cohen, N. J. (1984). Human memory and amnesia. In G. Lynch, J. McGough and N. Weinberger (Eds.), *Neurobiology of learning and memory.* New York: Guilford Press.

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

Squire, L. R. & Shimamura, A. (1997). The neuropsychology of memory dysfunction and its assessment. In T. E. Fienberg and M. J. Farah (eds.), *Behavioural neurology and neuropsychiatry* (pp. 232-262). New York: McGraw-Hill.

Squire, L.R. & Zola-Morgan, S. (1988). Memory: Brain systems and behaviour. *Trends in Neuroscience, 11,*170-175.

Squire, L.R. & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253,* 1380-1386.

Tabachnick, B. G. & Fidell, L. S. (1996). *Using multivariate statistics* (3rd Edn.). New York: Harper Collins Publishers.

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

Trahan, D. E. & Larrabee, G. J. (1988). *Continuous Visual Memory Test.* Odessa, FL: Psychological Assessment Resources.

Trahan, D. E., Larrabee, G. J., Quintana, J. W., Goethe, K. E., & Willingham, A. C. (1989). Development and clinical validation of an expanded paired associate test with delayed recall. *The Clinical Neuropsychologist, 3,* 169-183.

Trahan, D. E., Quintana, J. W., Willingham, A.C. & Goethe, K. E. (1988). The Visual Reproduction subtest: Standardisation & 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.

Trenerry, M. R., Jack, C. R., Ivnik, R. J., Sharbrough, F. W., Cascino, G. D., Hirschorn, K. A., Marsh, W. R., Kelly, P. J., & Meyer, F. B. (1993). MRI hippocampal volumes and memory function before or after temporal lobectomy. *Neurology, 43,* 1800-1805.

Troster, 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 & Huntington's diseases with the logical memory & visual reproduction tests. *Journal of Clinical & Experimental Neuropsychology, 15,* 773-788.

Tulsky, D. S. & Ledbetter, M. F. (2000). Updating to the WAIS-111 and WMS-111: Considerations for research and clinical practice. *Psychological Assessment,*

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

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385-98.

Tulving, E. (1987). Multiple memory systems and consciousness. *Human Neurobiology, 6,* 67-80.

Tulving, E. (1991). Concepts in human memory. In L. R. Squire, N. M. Weinberger, G. Lynch and J. L. Mc Gaugh (Eds.), *Memory: Organisation and locus of change* (pp 3-32). New York: Oxford University Press.

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

Vanderplas, J. M., & Garvin, E. A. (1959). The association value of random shapes. *Journal of Experimental Psychology, 57,* 147-154.

Warrington, E. K. (1984). *Manual for recognition memory test.* Windsor, England: NFER-Nelson.

Warrington, E. K. & James, M. (1967) An experimental investigation of facial recognition in patients with unilateral cerebral lesions, *Neuropsychologia, 5* , 253-266.

Watkins, M. J. & Gardiner, J. M. (1982). Cued recall. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 173-195). New York: Academic Press.

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

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

Wechsler, D. (1997). *Wechsler Adult Intelligence Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Memory Scale- Third Edition.* San Antonio, TX: The Psychological Corporation.

Weiskrantz, L. (1985). On issues and theories of the human amnesic syndrome. In N. M. Weinberger, J. L. McGough, and G. Lynch (Eds.), *Memory Systems of the brain: Animal and human cognitive processes.* New York: Guilford Press.

Williams, J. M. (1991). *Memory Assessment Scales Professional Manual.* Odessa: Psychology Assessment Resources.

Wilson, B. A., Cockburn, J., & Baddeley, A. (1985). The Rivermead Behavioural Memory Test. Reading, England: Tames Valley Test Co; Gaylord, MI: National Rehabilitation Services.

Woloszyn, D. B., Murphy, S. G., & Wetzel, L., & Fisher, W. (1993). Interrater agreement on the Wechsler Memory Scale- Revised in a mixed clinical population. *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.

Woodard, J. L. (1993). Confirmatory factor analysis of the Wechsler Memory Scale-Revised in a mixed clinical population. *Journal of Clinical & Experimental Neuropsychology, 15*, 968-973.

Zielinski, J. J. (1993). A comparison of the Wechsler Memory Scale-Revised and the Memory Assessment Scales: Administrative, clinical, & interpretive issues. *Professional Psychology Research & Practice, 24,* 353-359.

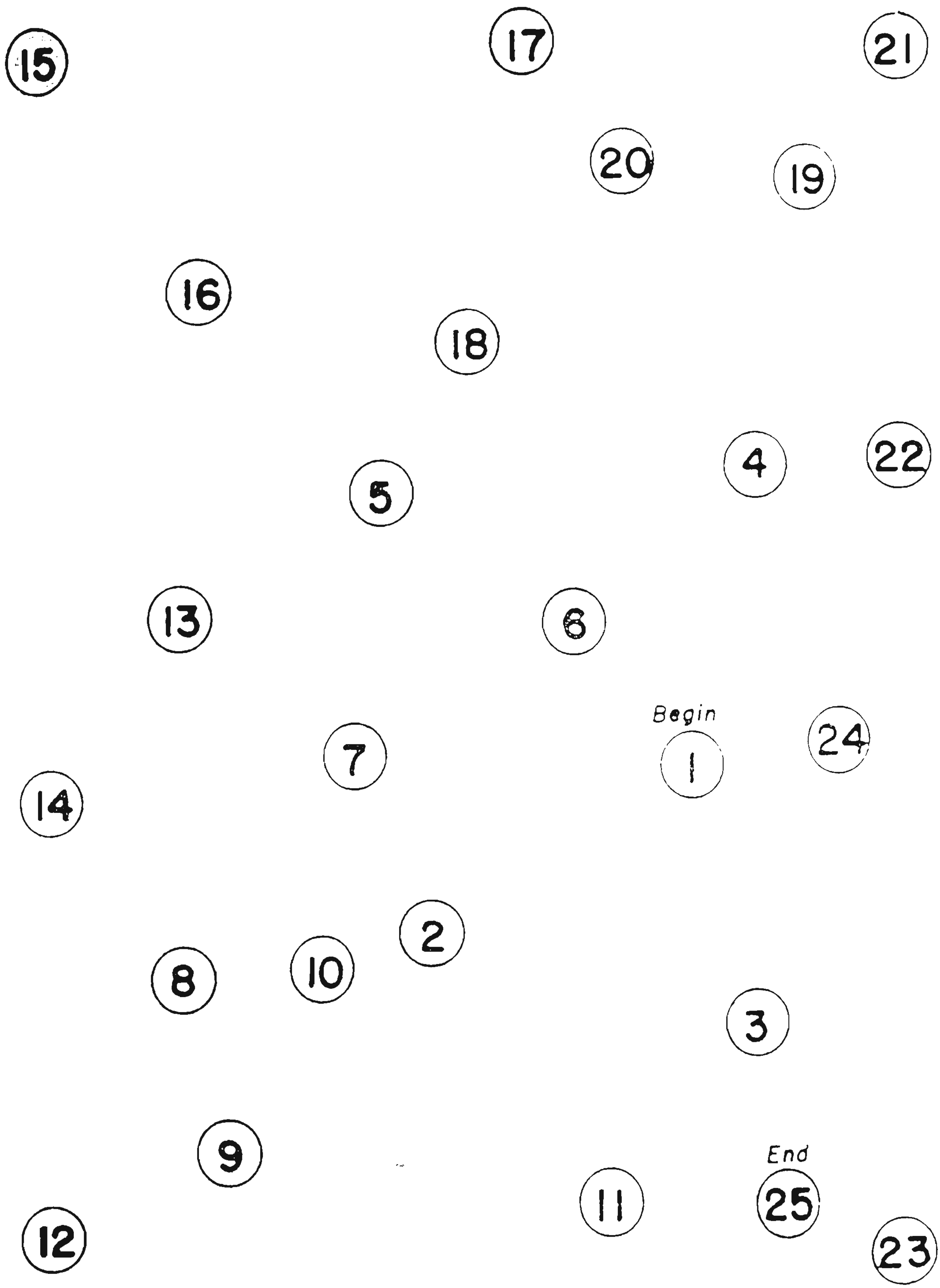
Zola, S. (1997). Amnesia: Neuroanatomic and clinical aspects. In T. E. Feinberg and M. J. Farah (Eds.), *Behavioural neurology and neuropsychology* (pp. 447-461). New York: McGraw Hill.

Zola-Morgan, S, Squire, L. R. & Amaral, D. G. (1989). Lesions of the hippocampal formation but not lesions of the fornix or the mamillary nuclei produce long-lasting memory impairment in monkeys. *Journal of Neuroscicene, 9,* 898-913.

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

APPENDIX A

Alternative Form for the Trail Making Test



APPENDIX B

PLAIN LANGUAGE STATEMENT

VICTORIA UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF PSYCHOLOGY

TITLE: MEMORY STUDY

My name is Therese Clark and I am undertaking a Doctor of Psychology program at Victoria University. As a part of the program I am carrying out a study of memory. The study is supervised by Dr. Peter Dowling, Lecturer in the Department of Psychology, Victoria University, St Albans.

I wish to study performance on some commonly used memory tasks with a view to gaining more information from them about memory skills. This may help us better understand the memory problems that some people report. I am particularly interested in memory functioning in people aged between 50 and 90.

The study will involve doing a range of tasks, most of which involve some aspect of memory.

We would welcome your involvement in this study. If you do chose to participate you are free to withdraw at any time. You are free to ask questions about any aspect of the study and your agreement to participate will be obtained in writing.

I will collect from you some general information like years of education, main occupation and medical history. The information I obtain from you will be confidential at all times.

The study will consist of two sessions lasting for approximately 45 minutes. Sessions can be conducted concurrently or on separate occasions. These sessions can be conducted in your home or at Victoria University. The sessions will be arranged at a convenient time for you.

Should you have any queries you can contact me through 9365 2353 or you may wish to call Dr. Peter Dowling on 9365 2556.

Thank you for your time

Therese Clark
Researcher

Dr. Peter Dowling
Clinical Neuropsychologist

APPENDIX C

Participant Consent Form

VICTORIA UNIVERSITY OF TECHNOLOGY

DEPARTMENT OF PSYCHOLOGY

CONSENT TO PARTICIPATE IN A RESEARCH PROJECT

TITLE OF RESEARCH PROJECT: Development of an Alternative Scoring System for the Visual Reproduction (VR) subtest of the Weschler Memory Scale-Revised (WMS-R).

RESEARCHER

I, certify that I have fully explained the aims, risks and procedures of the research to the participant named herein (or to the lawful guardian of such participant) and have handed to the participant (or guardian) a copy of this consent form together with an explanatory statement of the aims and procedures of the study and any risks to the participant.

I have explained to the participant that he/she is able to withdraw his/her consent at any time without consequence or penalty.

I undertake to the participant (or guardian) that the confidentiality and anonymity of the participant and his or her records will be preserved at all times.

In my opinion the participant (or guardian) appears to understand and wishes to participate.

Signed **Date.**

CONSENT OF THE PARTICIPANT OR GUARDIAN

The purpose of the above projects has been fully explained to me and I have read and signed the attached explanatory statement. I understand the aims and procedures of the study. I am aware that I can withdraw at any time from the study.

Participant's Name	Signature	Date
Witness's Name	Signature	Date

APPENDIX D

Skewness and Kurtosis Data

D.1: Skewness and Kurtosis Data for the Initial Comparison of Visual Reproduction Scoring Systems.

D.2: Skewness and Kurtosis Data for the Comparison of the Scoring Systems after Further Revision.

D.3: Skewness and Kurtosis Data for the Memory and Cognitive Measures- Experimental Group.

D.4: Skewness and Kurtosis Data for the Memory and Cognitive Measures - Control Group.

D.5: Skewness and Kurtosis Data on the Visual Reproduction Subtest- Left Hemisphere Lesion Group.

D.6: Skewness and Kurtosis Data on the Visual Reproduction Subtest- Right Hemisphere Lesion Group.

D.1: Skewness and Kurtosis Data for the Initial Comparison of Visual Reproduction Scoring Systems

	N	Skewness		Kurtosis	
Variable		Statistic	Std. Error	Statistic	Std. Error
Revised Scoring					
Design 1	44	-2.141	.357	4.779	.702
Design 2	44	-1.525	.357	4.244	.702
Design 3	44	-.815	.357	.348	.702
Design 4	44	-1.463	.357	1.941	.702
Total Score	44	-.830	.357	.731	.702
Original Scoring					
Design 1	44	-1.166	.357	1.820	.702
Design 2	44	-.669	.357	.601	.702
Design 3	44	.492	.357	-.528	.702
Design 4	44	-.544	.357	-.461	.702
Total Score	44	.061	.357	-.240	.702

D.2: Skewness and Kurtosis Data for the Comparison of the Scoring Systems After Further Revision

	N	Skewness		Kurtosis	
Design		Statistic	Std. Error	Statistic	Std. Error
Revised Scoring Immediate Recall					
Design 1	50	-.971	.337	.618	.662
Design 2	50	-.347	.337	-.597	.662
Design 3	50	-.831	.337	-.119	.662
Design 4	50	-1.594	.337	2.848	.662
Total Score	50	-.485	.337	-.736	.662
Revised Scoring Delayed Recall					
Design 1	50	-2.214	.337	3.841	.662
Design 2	50	-1.830	.337	3.361	.662
Design 3	50	-.996	.337	.484	.662
Design 4	50	-.834	.337	-.421	.662
Total Score	50	-.688	.337	-.795	.662
Original Scoring Immediate Recall					
Design 1	50	-.875	.337	1.270	.662
Design 2	50	-.385	.337	-1.109	.662
Design 3	50	-.751	.337	-.625	.662
Design 4	50	-1.163	.337	.502	.662
Total Score	50	-1.169	.337	.878	.662
Original Scoring Delayed Recall					
Design 1	50	-.994	.337	.654	.662
Design 2	50	-.987	.337	.712	.662
Design 3	50	-.534	.337	-.578	.662
Design 4	50	-.518	.337	-1.326	.662
Total Score	50	-.515	.337	-.933	.662

D.3: Skewness and Kurtosis Data on the Memory and Cognitive Measures-

Experimental Group

	N	Skewness		Kurtosis	
Variable		Statistic	Std. Error	Statistic	Std. Error
Visual Reproduction					
Revised Scoring Immediate Recall					
Design 1	30	-1.047	.427	.245	.833
Design 2	30	.138	.427	-.680	.833
Design 3	30	.307	.427	-.494	.833
Design 4	30	-.117	.427	-.626	.833
Total Score	30	-.126	.427	.029	.833
Revised Scoring Delayed Recall					
Design 1	30	.191	.427	-2.005	.833
Design 2	30	1.112	.427	-.506	.833
Design 3	30	-.038	.427	-1.099	.833
Design 4	30	.163	.427	-1.225	.833
Total Score	30	.022	.427	-.997	.833
Original Scoring Immediate Recall					
Design 1	30	-.619	.427	.303	.833
Design 2	30	-.266	.427	-.755	.833
Design 3	30	.552	.427	-.660	.833
Design 4	30	.316	.427	-1.040	.833
Total Score	30	.593	.427	-.022	.833
Original Scoring Delayed Recall					
Design 1	30	.531	.427	-1.405	.833
Design 2	30	1.195	.427	-.121	.833
Design 3	30	.364	.427	-1.167	.833
Design 4	30	.722	.427	-.869	.833
Total Score	30	.338	.427	-1.151	.833
Multiple Choice	30	-1.450	.427	.867	.833
Perceptual Match	30	-1.541	.427	2.294	.833
Logical Memory					
Immediate	30	.359	.427	.042	.833
Delayed	30	.532	.427	-.535	.833

(Appendix Continues on Next Page)

	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Verbal Paired Associates					
Immediate Associated	30	-1.048	.427	.280	.833
Immediate Novel	30	.576	.427	-.938	.833
Delayed Associated	30	-1.423	.427	1.454	.833
Delayed Novel	30	.325	.427	-1.336	.833
Faces					
Immediate Raw Score	30	-.100	.427	-.964	.833
Delayed Raw Score	30	.375	.427	-.190	.833
Block Design					
Raw Score	28	.538	.441	-1.015	.858
Scaled Score	28	.547	.441	-.760	.858
Vocabulary					
Raw Score	27	-.650	.448	.341	.872
Scaled Score	28	-.701	.441	1.928	.858

D.4: Skewness and Kurtosis Data on the Memory and Cognitive Measures-

Control Group

	N	Skewness		Kurtosis	
Variable		Statistic	Std. Error	Statistic	Std. Error
Visual Reproduction					
Revised Scoring Immediate Recall					
Design 1	30	-.964	.427	.130	.833
Design 2	30	-.137	.427	-.670	.833
Design 3	30	-.374	.427	-.314	.833
Design 4	30	-2.475	.427	7.260	.833
Total Score	30	-.929	.427	.364	.833
Revised Scoring Delayed Recall					
Design 1	30	-.451	.427	-1.789	.833
Design 2	30	-1.341	.427	1.756	.833
Design 3	30	-.230	.427	-.746	.833
Design 4	30	-.757	.427	-.629	.833
Total Score	30	-.530	.427	-.018	.833
Original Scoring Immediate Recall					
Design 1	30	-.356	.427	-.890	.833
Design 2	30	-.353	.427	-.661	.833
Design 3	30	-.546	.427	-.585	.833
Design 4	30	-1.103	.427	1.315	.833
Total Score	30	-.842	.427	.272	.833
Original Scoring Delayed Recall					
Design 1	30	-.206	.427	-1.793	.833
Design 2	30	-.537	.427	-.085	.833
Design 3	30	-.159	.427	-1.222	.833
Design 4	30	-.108	.427	-1.105	.833
Total Score	30	-.149	.427	-.618	.833
Multiple Choice	30	-4.782	.427	23.774	.833
Perceptual Match	30	-3.660	.427	12.207	.833
Logical Memory					
Immediate	30	-.270	.427	-.410	.833
Delayed	30	-.068	.427	-.459	.833

(Appendix Continues on Next Page)

	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Verbal Paired Associates					
Immediate Associated	30	-1.172	.427	.017	.833
Immediate Novel	30	-.073	.427	-1.249	.833
Delayed Associated	30	-2.428	.427	5.036	.833
Delayed Novel	30	.047	.427	-1.226	.833
Faces					
Immediate Raw Score	30	-.236	.427	.242	.833
Delayed Raw Score	30	.180	.427	-1.048	.833
Block Design					
Raw Score	30	.143	.427	-.599	.833
Scaled Score	30	.171	.427	.166	.833
Vocabulary					
Raw Score	30	-.023	.427	.161	.833
Scaled Score	30	.606	.427	.844	.833

D.5: Skewness and Kurtosis Data on the Visual Reproduction Subtest-

Left Hemisphere Lesion Group

	n	Skewness		Kurtosis	
Variable		Statistic	Std. Error	Statistic	Std. Error
Revised Scoring Immediate Recall					
Design 1	14	-.250	.597	-1.407	1.154
Design 2	14	.514	.597	-.786	1.154
Design 3	14	-.477	.597	-.449	1.154
Design 4	14	-2.190	.597	6.167	1.154
Total Score	14	.160	.597	-.606	1.154
Revised Scoring Delayed Recall					
Design 1	14	-.562	.597	-1.819	1.154
Design 2	14	.210	.597	-2.102	1.154
Design 3	14	-.605	.597	.065	1.154
Design 4	14	-1.324	.597	.827	1.154
Total Score	14	.361	.597	-.263	1.154
Original Scoring Immediate Recall					
Design 1	14	-1.157	.597	2.703	1.154
Design 2	14	.443	.597	.425	1.154
Design 3	14	-.087	.597	-1.762	1.154
Design 4	14	-1.127	.597	1.270	1.154
Total Score	14	.017	.597	-1.030	1.154
Original Scoring Delayed Recall					
Design 1	14	-.183	.597	-1.878	1.154
Design 2	14	.509	.597	-1.552	1.154
Design 3	14	-.190	.597	-1.506	1.154
Design 4	14	-1.006	.597	-.419	1.154
Total Score	14	-.298	.597	-.851	1.154

D.6: Skewness and Kurtosis Data on the Visual Reproduction Subtest-

Right Hemisphere Lesion Group

	n	Skewness		Kurtosis	
Variable		Statistic	Std. Error	Statistic	Std. Error
Revised Scoring Immediate Recall					
Design 1	16	-1.898	.564	4.278	1.091
Design 2	16	-.685	.564	.379	1.091
Design 3	16	-.981	.564	-.155	1.091
Design 4	16	-.307	.564	-.964	1.091
Total Score	16	-1.577	.564	2.613	1.091
Revised Scoring Delayed Recall					
Design 1	16	1.816	.564	1.547	1.091
Design 2	16	1.020	.564	-.577	1.091
Design 3	16	.224	.564	-1.392	1.091
Design 4	16	.469	.564	-.704	1.091
Total Score	16	.282	.564	-.741	1.091
Original Scoring Immediate Recall					
Design 1	16	-1.177	.564	1.434	1.091
Design 2	16	-.395	.564	-.403	1.091
Design 3	16	.295	.564	-1.522	1.091
Design 4	16	.306	.564	-1.446	1.091
Total Score	16	-.573	.564	-.111	1.091
Original Scoring Delayed Recall					
Design 1	16	1.988	.564	2.698	1.091
Design 2	16	1.373	.564	.743	1.091
Design 3	16	.537	.564	-.964	1.091
Design 4	16	1.838	.564	2.712	1.091
Total Score	16	-.084	.564	-1.458	1.091

APPENDIX E

Normality Tests

E.1: Normality Test Results for the Initial Comparison of Visual Reproduction Scoring Systems.

E.2: Normality Test Results for the Comparison of the Scoring Systems after Further Revision.

E.3: Normality Test Results for Memory and Cognitive Measures- Experimental Group.

E.4: Normality Test Results for Memory and Cognitive Measures- Control Group.

E.5: Normality Test Results for Visual Reproduction- Left Hemisphere Lesion Group.

E.6: Normality Test Results for Visual Reproduction - Right Hemisphere Lesion Group.

E.1: Normality Test Results for Initial Comparison of Visual Reproduction Scoring Systems

	Kolmogorov-Smirnov		
Variable.	Statistic	df	*p*
Revised Scoring			
Design 1	.245	44	.000
Design 2	.159	44	.007
Design 3	.145	44	.021
Design 4	.208	44	.000
Total Score	.106	44	.200
Original Scoring			
Design 1	.271	44	.000
Design 2	.170	44	.003
Design 3	.248	44	.000
Design 4	.151	44	.014
Total Score	.088	44	.200

E.2: Normality Test Results for Comparison of the Scoring Systems after Further Revision

	Kolmogorov-Smirnov		
Variable	Statistic	df	*p*
Revised Scoring Immediate Recall			
Design 1	.227	50	.000
Design 2	.154	50	.005
Design 3	.172	50	.001
Design 4	.181	50	.000
Total Score	.137	50	.020
Revised Scoring Delayed Recall			
Design 1	.268	50	.000
Design 2	.195	50	.000
Design 3	.197	50	.000
Design 4	.208	50	.000
Total Score	.178	50	.000
Original Scoring Immediate Recall			
Design 1	.194	50	.000
Design 2	.207	50	.000
Design 3	.213	50	.000
Design 4	.169	50	.001
Total Score	.154	50	.004
Original Scoring Delayed Recall			
Design 1	.193	50	.000
Design 2	.198	50	.000
Design 3	.160	50	.003
Design 4	.206	50	.000
Total Score	.142	50	.013

E.3: Normality Test Results for Memory and Cognitive Measures-

Experimental Group

	Kolmogorov-Smirnov		
Variable	Statistic	df	*p*
Visual Reproduction			
Revised Scoring Immediate Recall			
Design 1	.239	30	.000
Design 2	.119	30	.200
Design 3	.158	30	.054
Design 4	.121	30	.200
Total Score	.119	30	.200
Revised Scoring Delayed Recall			
Design 1	.324	30	.000
Design 2	.378	30	.000
Design 3	.147	30	.097
Design 4	.239	30	.000
Total Score	.099	30	.200
Original Scoring Immediate Recall			
Design 1	.154	30	.066
Design 2	.161	30	.045
Design 3	.217	30	.001
Design 4	.121	30	.200
Total Score	.150	30	.082
Original Scoring Delayed Recall			
Design 1	.337	30	.000
Design 2	.406	30	.000
Design 3	.200	30	.004
Design 4	.215	30	.001
Total Score	.110	30	.200
Multiple Choice	.396	30	.000
Perceptual Match	.354	30	.000
Logical Memory			
Immediate Recall Accuracy	.082	30	.200
Delayed Recall Accuracy	.102	30	.200

(Appendix Continues on Next Page)

	Kolmogorov-Smirnov		
Variable	Statistic	df	*p*
Verbal Paired Associates			
Immediate Recall Associated Pairs	.218	30	.001
Immediate Recall Novel Pairs	.154	30	.068
Delayed Recall Associated Pairs	.267	30	.000
Delayed Recall Novel Pairs	.222	30	.001
Faces			
Raw Score Immediate Recognition	.106	30	.200
Raw Score Delayed Recognition	.117	30	.200
Block Design			
Raw Score	.136	28	.199
Scaled Score	.183	28	.018
Vocabulary			
Raw Score	.129	27	.200
Scaled Score	.178	27	.027

E.4: Normality Test Results for Memory and Cognitive Measures-

Control Group

	Kolmogorov-Smirnov		
Variable	Statistic	df	*p*
Visual Reproduction			
Revised Scoring Immediate Recall			
Design 1	.204	30	.003
Design 2	.125	30	.200
Design 3	.129	30	.200
Design 4	.276	30	.000
Total Score	.214	30	.001
Revised Scoring Delayed Recall			
Design 1	.263	30	.000
Design 2	.161	30	.047
Design 3	.107	30	.200
Design 4	.206	30	.002
Total Score	.088	30	.200
Original Scoring Immediate Recall			
Design 1	.268	30	.000
Design 2	.185	30	.011
Design 3	.202	30	.003
Design 4	.182	30	.012
Total Score	.126	30	.200
Original Scoring Delayed Recall			
Design 1	.250	30	.000
Design 2	.164	30	.039
Design 3	.167	30	.031
Design 4	.104	30	.200
Total Score	.116	30	.200
Multiple Choice	.526	30	.000
Perceptual Match	.537	30	.000
Logical Memory			
Immediate Recall Accuracy	.114	30	.200
Delayed Recall Accuracy	.107	30	.200

(Appendix Continues on Next Page)

	Kolmogorov-Smirnov		
Variable	Statistic	df	*p*
Verbal Paired Associates			
Immediate Recall Associated Pairs	.356	30	.000
Immediate Recall Novel Pairs	.141	30	.133
Delayed Recall Associated Pairs	.493	30	.000
Delayed Recall Novel Pairs	.183	30	.012
Faces			
Raw Score Immediate Recognition	.116	30	.200
Raw Score Delayed Recognition	.186	30	.010
Block Design			
Raw Score	.113	30	.200
Scaled Score	.203	30	.003
Vocabulary			
Raw Score	.101	30	.200
Scaled Score	.171	30	.025

E.5: Normality Test Results for Memory Measures-

Left Hemisphere Lesion Group

	Kolmogorov-Smirnov		
Variable	Statistic	df	*p*
Revised Scoring Immediate Recall			
Design 1	.159	14	.200
Design 2	.191	14	.175
Design 3	.151	14	.200
Design 4	.196	14	.151
Total Score	.214	14	.080
Revised Scoring Delayed Recall			
Design 1	.279	14	.004
Design 2	.323	14	.000
Design 3	.114	14	.200
Design 4	.248	14	.019
Total Score	.147	14	.200
Original Scoring Immediate Recall			
Design 1	.268	14	.007
Design 2	.182	14	.200
Design 3	.233	14	.037
Design 4	.183	14	.200
Total Score	.137	14	.200
Original Scoring Delayed Recall			
Design 1	.237	14	.032
Design 2	.308	14	.001
Design 3	.241	14	.027
Design 4	.193	14	.168
Total Score	.132	14	.200

E.6: Normality Test Results for Memory Measures-

Right Hemisphere Lesion Group

	Kolmogorov-Smirnov		
Variable	Statistic	df	*p*
Revised Scoring Immediate Recall			
Design 1	.193	16	.113
Design 2	.157	16	.200
Design 3	.230	16	.024
Design 4	.214	16	.049
Total Score	.206	16	.067
Revised Scoring Delayed Recall			
Design 1	.491	16	.000
Design 2	.380	16	.000
Design 3	.283	16	.001
Design 4	.276	16	.002
Total Score	.103	16	.200
Original Scoring Immediate Recall			
Design 1	.290	16	.001
Design 2	.145	16	.200
Design 3	.217	16	.043
Design 4	.168	16	.200
Total Score	.152	16	.200
Original Scoring Delayed Recall			
Design 1	.488	16	.000
Design 2	.367	16	.000
Design 3	.315	16	.000
Design 4	.296	16	.001
Total Score	.179	16	.178

APPENDIX F

Intraclass Correlations- Intra-rater Reliability

Immediate Recall

	Intraclass Correlations		
Design	Original Criteria	Revised Criteria	*N*
1	.91	.99	60
2	.88	.97	60
3	.95	.99	60
4	.95	.99	60
Total	.97	.99	60

Delayed recall

	Intraclass Correlation			
Design	Original Criteria	*n*	Revised Criteria	*n*
1	.88	33	.97	34
2	.86	37	.97	38
3	.98	49	.99	53
4	.97	43	.98	44
Total	.99	54	.99	57

APPENDIX G

Individual Item Agreement Across Two Scoring Occasions

Original Scoring System

Item	Design 1		Design 2		Design 3		Design 4	
	Imm	Delay	Imm	Delay	Imm	Delay	Imm	Delay
1	59	59	59	60	59	59	59	60
2	52	57	58	59	59	59	56	59
3	56	59	59	58	59	60	53	59
4	59	55	58	58	56	56	55	59
5	59	58	60	56	59	59	58	58
6	57	58	51	55	53	58	57	57
7	53	54	50	53	56	60	57	58
8					57	60	55	56
9					53	59	57	58
10							56	60
11							60	60
12							58	60
13							50	56
14							57	60
15							57	59
16							58	55
17							57	58
18							58	55

(Appendix Continues on Next Page)

Revised Scoring System

Individual Item Agreement Across Two Scoring Occasions

Item	Design 1		Design 2		Design 3		Design 4	
	Imm	Delay	Imm	Delay	Imm	Delay	Imm	Delay
1	60	60	60	60	60	60	60	57
2	60	59	60	60	60	59	60	59
3	59	57	60	59	59	60	60	59
4	60	59	60	60	50	55	58	60
5	57	60	60	60	60	59	55	58
6	60	60	60	60	59	60	58	60
7	60	60	60	60	59	60	58	59
8	60	59	60	60	60	60	60	59
9	54	57	60	57	59	60	59	58
10	53	58	59	60	60	60	60	60
11	60	60	60	60	60	60	60	60
12	60	59	56	55	60	60	60	60
13	60	60	54	55	57	56	56	57
14	60	60	55	55	58	59	60	58
15	60	60	53	58	59	60	60	60
16	57	60	59	60	59	60	60	60
17	58	60	54	58	59	60	59	60
18	59	59	55	58	59	59	58	58
19	59	58	57	56	58	56	59	59
20	58	59	60	60	59	58	59	60

APPENDIX H

Intraclass Correlations- Inter-rater Reliability

Immediate Recall

	Intra class Correlations		
Design	Original Criteria	Revised Criteria	*N*
1	.81	.97	30
2	.88	.91	30
3	.82	.96	30
4	.94	.98	30
Total	.95	.98	30

Delayed Recall

	Intra class Correlation			
Design	Original Criteria	*n*	Revised Criteria	*n*
1	.69	23	.96	25
2	.88	23	.96	24
3	.90	24	.91	27
4	.92	24	.93	24
Total	.98	29	.99	29

APPENDIX I

Individual Item Agreement Across Scorers

<u>Original Scoring System</u>

Item	Design 1		Design 2		Design 3		Design 4	
	Imm	Delay	Imm	Delay	Imm	Delay	Imm	Delay
1	27	28	30	29	25	24	28	28
2	28	28	28	29	27	28	24	26
3	20	26	27	27	27	27	25	27
4	24	26	28	27	23	25	27	28
5	30	30	28	28	26	28	28	26
6	28	28	26	28	6	26	25	27
7	23	27	22	23	27	29	25	26
8					24	28	20	28
9					26	30	26	28
10							27	27
11							26	23
12							29	30
13							22	26
14							25	25
15							28	29
16							27	28
17							27	28
18							29	29

(Appendix Continued on Next Page)

<u>Revised Scoring System</u>

Individual Item Agreement Across Scorers

Item	Design 1		Design 2		Design 3		Design 4	
	Imm	Delay	Imm	Delay	Imm	Delay	Imm	Delay
1	30	30	30	30	30	28	30	25
2	28	27	30	30	30	28	26	28
3	30	27	30	29	25	25	29	29
4	30	30	30	30	26	29	29	28
5	28	6	30	30	28	28	24	29
6	30	30	30	30	29	29	30	30
7	30	29	29	30	30	29	26	29
8	28	27	26	28	29	26	28	27
9	28	29	27	25	29	29	30	28
10	23	27	27	29	29	30	30	29
11	29	27	28	29	30	25	29	30
12	30	29	26	28	29	28	27	30
13	29	28	25	30	16	13	26	28
14	30	30	26	27	23	30	29	27
15	30	27	24	26	29	29	30	30
16	26	27	28	29	30	30	26	26
17	29	28	26	26	28	29	30	30
18	30	29	25	22	28	28	28	29
19	29	29	26	25	30	30	28	26
20	28	28	30	27	28	28	28	26

APPENDIX J

Subject Information- Experimental and Control Groups

ID	Age	Sex	Years of Education	Employment	Group
1	80	Female	9	<10	Experimental
2	54	Male	13	10-15	Experimental
3	76	Male	9	<10	Experimental
4	58	Female	9	<10	Experimental
5	64	Male	9	<10	Experimental
6	84	Male	13	10-15	Experimental
7	83	Female	13	10-15	Experimental
8	58	Female	9	<10	Experimental
9	60	Male	9	<10	Experimental
10	67	Female	9	<10	Control
11	77	Male	11	10-15	Experimental
12	72	Male	9	<10	Control
13	87	Female	9	<10	Experimental
14	59	Male	10	<10	Experimental
15	75	Male	7	<10	Experimental
16	70	Female	9	<10	Experimental
17	80	Female	9	<10	Experimental
18	79	Male	10	<10	Experimental
19	85	Female	9	<10	Control
20	83	Female	9	<10	Control
21	64	Female	9	<10	Control
22	74	Female	9	<10	Experimental
23	74	Female	13	10-15	Experimental
24	61	Male	9	<10	Experimental
25	70	Male	9	<10	Control
26	79	Female	9	<10	Control
27	74	Female	9	<10	Experimental
28	76	Male	9	<10	Experimental
29	68	Male	8	<10	Experimental
30	70	Female	9	<10	Control
31	68	Male	10	<10	Control
32	67	Female	9	<10	Experimental
33	87	Male	9	<10	Experimental
34	64	Female	9	<10	Experimental
35	60	Male	11	10-15	Control
36	55	Male	11	10-15	Control
37	56	Female	10	<10	Control

(Appendix Continues on Next Page)

ID	Age	Sex	Years of Education	Employment	Group
38	50	Male	11	10-15	Control
39	71	Male	11	<10	Control
40	68	Male	9	<<10	Experimental
41	52	Male	10	10-15	Control
42	55	Male	10	10-15	Control
43	53	Female	11	<10	Control
44	53	Male	8	<10	Control
45	67	Female	8	<10	Experimental
46	67	Female	8	<10	Control
47	82	Female	8	<10	Control
48	84	Female	8	<10	Experimental
49	54	Female	11	<10	Experimental
50	58	Male	7	<10	Experimental
51	76	Male	9	<10	Control
52	68	Female	14	10-15	Control
53	76	Female	8	<10	Control
54	78	Male	9	<10	Control
55	75	Female	9	<10	Control
56	76	Female	11	10-15	Control
57	82	Female	8	<10	Control
58	71	Female	9	<10	Control
59	73	Male	10	10-15	Control
60	78	Female	9	<10	Control

APPENDIX K

Spearman's Correlations

K.1: Correlation between immediate recall subtests for the control group and the experimental group

K.2: Correlation between delayed recall subtests for the control group and the experimental group

K.1: Correlation Between Immediate Recall Subtests.

Control Group

		1	2	3	4	5	6
1	VPA Familiar Pairs						
2	VPA Novel Pairs	.60					
3	Logical Memory	.08	.27				
4	Faces	.34	.34	.29			
5	Revised VR	.38	.56	.37	.39		
6	Original VR	.29	.48	.38	.39	.89	

Experimental Group

		1	2	3	4	5	6
1	VPA Familiar Pairs						
2	VPA Novel Pairs	.51					
3	Logical Memory	.61	.60				
4	Faces	.05	.53	.42			
5	Revised VR	.34	.47	.27	.35		
6	Original VR	.27	.36	.23	.26	.88	

K.2: Correlation between Delayed Recall Subtests.

Control Group

		1	2	3	4	5	6
1	VPA Familiar Pairs						
2	VPA Novel Pairs	.46					
3	Logical Memory	.13	.25				
4	Faces	.01	.30	.34			
5	Revised VR	.16	.43	.24	.58		
6	Original VR	.34	.48	.21	.44	.93	

Experimental Group

		1	2	3	4	5	6
1	VPA Familiar Pairs						
2	VPA Novel Pairs	.74					
3	Logical Memory	.61	.62				
4	Faces	.25	.42	.44			
5	Revised VR	.51	.42	.58	.44		
6	Original VR	.54	.47	.55	.48	.99	

APPENDIX L

Mean Ranks and Sum of Ranks for the Experimental and Control Groups on Memory and Cognitive Measures

	Group Membership	N	Mean Rank	Sum of Ranks
Visual Reproduction				
Revised Scoring Immediate Recall				
Design 1	Experimental	30	26.65	799.50
	Control	30	34.35	1030.50
Design 2	Experimental	30	25.47	764.00
	Control	30	35.53	1066.00
Design 3	Experimental	30	22.02	660.50
	Control	30	38.98	1169.50
Design 4	Experimental	30	21.75	652.50
	Control	30	39.25	1177.50
Total Score	Experimental	30	21.98	659.50
	Control	30	39.02	1170.50
Revised Scoring Delayed Recall				
Design 1	Experimental	30	25.63	769.00
	Control	30	35.37	1061.00
Design 2	Experimental	30	21.55	646.50
	Control	30	39.45	1183.50
Design 3	Experimental	30	24.10	723.00
	Control	30	36.90	1107.00
Design 4	Experimental	30	24.90	747.00
	Control	30	36.10	1083.00
Total Score	Experimental	30	21.85	655.50
	Control	30	39.15	1174.50
Original Scoring Immediate Recall				
Design 1	Experimental	30	28.00	840.00
	Control	30	33.00	990.00
Design 2	Experimental	30	27.38	821.50
	Control	30	33.62	1008.50
Design 3	Experimental	30	22.35	670.50
	Control	30	38.65	1159.50
Design 4	Experimental	30	22.18	665.50
	Control	30	38.82	1164.50
Total	Experimental	30	21.67	650.00
	Control	30	39.33	1180.00

(Appendix Continues on Next Page)

Original Scoring Delayed Recall				
Design 1	Experimental	30	25.52	765.50
	Control	30	35.48	1064.50
Design 2	Experimental	30	22.35	670.50
	Control	30	38.65	1159.50
Design 3	Experimental	30	24.83	745.00
	Control	30	36.17	1085.00
Design 4	Experimental	30	25.20	756.00
	Control	30	35.80	1074.00
Total Score	Experimental	30	22.97	689.00
	Control	30	38.03	1141.00
Multiple Choice	Experimental	30	26.47	794.00
	Control	30	34.53	1036.00
Perceptual match	Experimental	30	25.40	762.00
	Control	30	35.60	1068.00
Block Design				
Raw score	Experimental	28	22.41	627.50
	Control	30	36.12	1083.50
Scaled Score	Experimental	28	21.66	606.50
	Control	30	36.82	1104.50
Vocabulary–R				
Raw score	Experimental	27	21.69	585.50
	Control	30	35.58	1067.50
Scaled score	Experimental	28	22.20	621.50
	Control	30	36.32	1089.50

APPENDIX M

Demographic Data for the Left and Right Lesion Groups

ID	AGE	SEX	GROUP
1	73	Female	Left
2	56	Male	Left
3	42	Male	Left
4	64	Female	Left
5	55	Male	Left
6	63	Male	Left
7	43	Female	Left
8	66	Female	Left
9	77	Female	Left
10	84	Male	Left
11	83	Female	Left
12	59	Male	Left
13	74	Female	Left
14	64	Female	Left
15	70	Male	Right
16	48	Female	Right
17	48	Male	Right
18	78	Female	Right
19	67	Female	Right
20	66	Male	Right
21	58	Female	Right
22	77	Male	Right
23	80	Female	Right
24	69	Female	Right
25	69	Male	Right
26	74	Female	Right
27	61	Male	Right
28	76	Male	Right
29	68	Male	Right
30	84	Female	Right

APPENDIX N

Mean Ranks and Sum of Ranks for the Left and Right Lesion Groups on the Visual Reproduction Subtest

Variable	Group Membership	N	Mean Rank	Sum of Ranks
Revised Scoring Immediate Recall				
Design 1	Left	14	17.43	244.00
	Right	16	13.81	221.00
Design 2	Left	14	17.25	241.50
	Right	16	13.97	223.50
Design 3	Left	14	20.43	286.00
	Right	16	11.19	179.00
Design 4	Left	14	20.18	282.50
	Right	16	11,41	182.50
Total	Left	14	20.32	284.50
	Right	16	11.28	180.50
Revised Scoring Delayed Recall				
Design 1	Left	14	19.50	273.00
	Right	16	12.00	192.00
Design 2	Left	14	17.14	240.00
	Right	16	14.06	225.00
Design 3	Left	14	20.64	289.00
	Right	16	11.00	176.00
Design 4	Left	14	20.71	290.00
	Right	16	10.94	175.00
Total	Left	14	22.61	316.50
	Right	16	9,28	148.50
Original Scoring Immediate Recall				
Design 1	Left	14	17.79	249.00
	Right	16	13.50	216.00
Design 2	Left	14	15.68	219.50
	Right	16	15.34	245.50
Design 3	Left	14	18.89	264.50
	Right	16	12.53	200.50
Design 4	Left	14	20.79	291.00
	Right	16	10.88	174.00
Total	Left	14	20.46	286.50
	Right	16	11.16	178.50

(Appendix Continues on Next Page)

Variable	Group Membership	N	Mean Rank	Sum of Ranks
Original Delayed Recall				
Design 1	Left	14	19.18	268.50
	Right	16	12.28	196.50
Design 2	Left	14	16.86	236.00
	Right	16	14.31	229.00
Design 3	Left	14	20.00	280.00
	Right	16	11.56	185.00
Design 4	Left	14	21.21	297.00
	Right	16	10.50	168.00
Total	Left	14	23.04	322.50
	Right	16	8.91	142.50

APPENDIX 0

O.1: Mann- Whitney U Results for Immediate Recall- Design 1 Revised Scoring

O.2: Mann- Whitney U Results for Immediate Recall- Design 2 Revised Scoring

O.3: Mann- Whitney U Results for Immediate Recall- Design 3 Revised Scoring

O.4: Mann- Whitney U Results for Immediate Recall- Design 4 Revised Scoring

O.5: Mann- Whitney U Results for Delayed Recall- Design 1 Revised Scoring

O.6: Mann- Whitney U Results for Delayed Recall- Design 2 Revised Scoring

O.7: Mann- Whitney U Results for Delayed Recall- Design 3 Revised Scoring

O.8: Mann- Whitney U Results for Delayed Recall- Design 4 Revised Scoring

O.9: Mann- Whitney U Results for Immediate Recall- Design 1 Original Scoring

O.10: Mann- Whitney U Results for Immediate Recall- Design 2 Original Scoring

O.11: Mann- Whitney U Results for Immediate Recall- Design 3 Original Scoring

O.12: Mann- Whitney U Results for Immediate Recall- Design 4 Original Scoring

O.13: Mann- Whitney U Results for Delayed Recall- Design 1 Original Scoring

O.14: Mann- Whitney U Results for Delayed Recall- Design 2 Original Scoring

O.15: Mann- Whitney U Results for Delayed Recall- Design 3 Original Scoring

O.16: Mann- Whitney U Results for Delayed Recall- Design 4 Original Scoring

O.1: Mann- Whitney U Test Results for Immediate Recall-

Design 1 Revised Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	105.000	241.000	-.935	.790
2	106.000	242.000	-.480	.822
3	91.000	227.000	-1.679	.400
4	106.000	242.000	-.480	.822
5	111.000	216.000	-.096	.984
6	99.000	235.000	-.917	.608
7	105.000	241.000	-.935	.790
8	92.000	228.000	-1.287	.423
9	106.000	211.000	-.480	.822
10	93.000	229.000	-1.031	.448
11	99.000	235.000	-.917	.608
12	98.000	234.000	-1.346	.580
13	105.000	241.000	-.935	.790
14	105.000	241.000	-.935	.790
15	98.000	234.000	-1.346	.580
16	92.000	228.000	-1.287	.423
17	102.000	238.000	-.509	.697
18	94.000	230.000	-.942	.473
19	85.000	221.000	-1.413	.275
20	84.000	220.000	-1.976	.257
Not corrected for ties.				

O.2: Mann- Whitney U Test Results for Immediate Recall-

Design 2 Revised Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	112.000	248.000	.000	1.000
2	105.000	241.000	-.935	.790
3	91.000	227.000	-1.679	.400
4	112.000	248.000	.000	1.000
5	105.000	241.000	-.935	.790
6	105.000	241.000	-.935	.790
7	107.000	243.000	-.322	.854
8	93.000	229.000	-1.078	.448
9	95.000	231.000	-.846	.498
10	75.000	211.000	-1.779	.131
11	93.000	229.000	-1.031	.448
12	108.000	244.000	-.227	.886
13	102.000	238.000	-.509	.697
14	94.000	230.000	-1.079	.473
15	83.000	219.000	-1.404	.240
16	85.000	221.000	-1.413	.275
17	84.000	189.000	-1.394	.257
18	75.000	180.000	-1.779	.131
19	109.000	214.000	-.157	.918
20	112.000	248.000	.000	1.000
Not corrected for ties.				

O.3: Mann- Whitney U Test Results for Immediate Recall-

Design 3 Revised Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	70.000	206.000	-2.519	.085
2	84.000	220.000	-1.976	.257
3	63.000	199.000	-2.779	.043
4	103.000	239.000	-.720	.728
5	75.000	211.000	-1.779	.131
6	68.000	204.000	-2.130	.070
7	61.000	197.000	-2.497	.034
8	101.000	237.000	-.597	.667
9	80.000	216.000	-1.593	.193
10	104.000	240.000	-.385	.759
11	90.000	226.000	-1.058	.377
12	98.000	234.000	-.678	.580
13	108.000	213.000	-.227	.886
14	95.000	231.000	-1.200	.498
15	97.000	233.000	-.720	.552
16	79.000	215.000	-1.727	.179
17	82.000	218.000	-1.439	.224
18	80.000	216.000	-1.593	.193
19	96.000	232.000	-1.539	.525
20	102.000	238.000	-.509	.697
Not corrected for ties.				

O.4: Mann- Whitney U Test Results for Immediate Recall-

Design 4 Revised Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	99.000	235.000	-.917	.608
2	73.000	209.000	-1.910	.110
3	71.000	207.000	-2.224	.093
4	73.000	209.000	-1.910	.110
5	108.000	213.000	-.227	.886
6	52.000	188.000	-2.878	.012
7	71.000	207.000	-2.325	.093
8	45.000	181.000	-3.221	.004
9	88.000	224.000	-1.919	.334
10	85.000	221.000	-1.619	.275
11	79.000	215.000	-1.727	.179
12	103.000	239.000	-.441	.728
13	76.000	212.000	-1.763	.142
14	85.000	221.000	-1.619	.275
15	100.000	236.000	-.720	.637
16	90.000	226.000	-1.058	.377
17	87.000	223.000	-1.272	.313
18	95.000	231.000	-.846	.498
19	83.000	219.000	-1.404	.240
20	95.000	231.000	-.846	.498
Not corrected for ties.				

O.5: Mann- Whitney U Test Results for Delayed Recall-

Design 1 Revised Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	61.000	197.000	-2.497	.034
2	69.000	205.000	-2.140	.077
3	77.000	213.000	-1.781	.154
4	62.000	198.000	-2.544	.038
5	62.000	198.000	-2.544	.038
6	77.000	213.000	-1.781	.154
7	61.000	197.000	-2.497	.034
8	77.000	213.000	-1.781	.154
9	104.000	240.000	-1.069	.759
10	95.000	231.000	-1.200	.498
11	77.000	213.000	-1.781	.154
12	62.000	198.000	-2.544	.038
13	61.000	197.000	-2.497	.034
14	61.000	197.000	-2.497	.034
15	54.000	190.000	-2.887	.015
16	62.000	198.000	-2.544	.038
17	70.000	206.000	-2.198	.085
18	69.000	205.000	-2.140	.077
19	96.000	232.000	-1.539	.525
20	54.000	190.000	-2.887	.015
Not corrected for ties.				

O.6 Mann- Whitney U Test Results for Delayed Recall-

Design 2 Revised Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	98.000	234.000	-.678	.580
2	98.000	234.000	-.678	.580
3	92.000	228.000	-1.018	.423
4	98.000	234.000	-.678	.580
5	105.000	241.000	-.337	.790
6	92.000	228.000	-1.018	.423
7	92.000	228.000	-1.018	.423
8	93.000	229.000	-1.031	.448
9	94.000	230.000	-1.079	.473
10	102.000	238.000	-.644	.697
11	102.000	238.000	-.644	.697
12	93.000	229.000	-1.031	.448
13	79.000	215.000	-1.979	.179
14	96.000	232.000	-1.539	.525
15	108.000	213.000	-.227	.886
16	103.000	239.000	-.720	.728
17	107.000	243.000	-.254	.854
18	92.000	228.000	-1.018	.423
19	88.000	224.000	-1.919	.334
20	91.000	227.000	-1.028	.400

Not corrected for ties.

O.7: Mann- Whitney U Test Results for Delayed Recall-

Design 3 Revised Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	64.000	200.000	-2.513	.047
2	72.000	208.000	-2.035	.101
3	65.000	201.000	-2.340	.052
4	102.000	238.000	-.644	.697
5	91.000	227.000	-1.028	.400
6	83.000	219.000	-1.404	.240
7	84.000	220.000	-1.394	.257
8	45.000	181.000	-3.221	.004
9	53.000	189.000	-2.856	.013
10	63.000	199.000	-2.658	.043
11	62.000	198.000	-2.544	.038
12	77.000	213.000	-1.781	.154
13	87.000	223.000	-1.609	.313
14	111.000	247.000	-.096	.984
15	63.000	199.000	-2.658	.043
16	65.000	201.000	-2.340	.052
17	69.000	205.000	-2.140	.077
18	73.000	209.000	-1.910	.110
19	104.000	240.000	-1.069	.759
20	80.000	216.000	-1.593	.193
Not corrected for ties.				

O.8: Mann- Whitney U Test Results for Delayed Recall-

Design 4 Revised Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	87.000	223.000	-1.272	.313
2	60.000	196.000	-2.500	.031
3	52.000	188.000	-2.878	.012
4	60.000	196.000	-2.500	.031
5	110.000	246.000	-.141	.951
6	45.000	181.000	-3.221	.004
7	94.000	230.000	-1.079	.473
8	47.000	183.000	-3.308	.006
9	88.000	224.000	-1.919	.334
10	80.000	216.000	-1.593	.193
11	74.000	210.000	-1.827	.120
12	91.000	227.000	-1.028	.400
13	84.000	220.000	-1.394	.257
14	68.000	204.000	-2.130	.070
15	73.000	209.000	-1.910	.110
16	69.000	205.000	-2.140	.077
17	53.000	189.000	-2.856	.013
18	61.000	197.000	-2.497	.034
19	84.000	220.000	-1.394	.257
20	89.000	225.000	-1.106	.355
Not corrected for ties.				

O. 9: Mann- Whitney U Test Results for Immediate Recall-

Design 1 Original Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	106.000	242.000	-.480	.822
2	85.000	221.000	-1.619	.275
3	110.000	215.000	-.141	.951
4	92.000	228.000	-1.287	.423
5	77.000	213.000	-1.781	.154
6	95.000	231.000	-.846	.498
7	104.000	209.000	-.385	.759
Not corrected for ties.				

O.10: Mann- Whitney U Test Results for Immediate Recall-

Design 2 Original Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	105.000	241.000	-.935	.790
2	98.000	234.000	-1.346	.580
3	100.000	236.000	-.720	.637
4	97.000	233.000	-.720	.552
5	79.000	215.000	-1.979	.179
6	73.000	178.000	-1.910	.110
7	93.000	198.000	-1.078	.448
Not corrected for ties.				

O.11: Mann- Whitney U Test Results for Immediate Recall-

Design 3 Original Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	77.000	213.000	-2.253	.154
2	68.000	204.000	-2.130	.070
3	70.000	206.000	-2.519	.085
4	110.000	215.000	-.100	.951
5	97.000	233.000	-.720	.552
6	97.000	233.000	-.720	.552
7	93.000	229.000	-1.031	.448
8	90.000	226.000	-1.058	.377
9	103.000	239.000	-.720	.728
Not corrected for ties.				

O.12: Mann- Whitney U Test Results for Immediate Recall-

Design 4 Original Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	81.000	217.000	-1.501	.208
2	72.000	208.000	-2.035	.101
3	31.000	167.000	-3.966	.000
4	52.000	188.000	-2.878	.012
5	88.000	224.000	-1.919	.334
6	80.000	216.000	-2.258	.193
7	53.000	189.000	-2.856	.013
8	85.000	221.000	-1.413	.275
9	60.000	196.000	-2.500	.031
10	38.000	174.000	-3.582	.001
11	78.000	214.000	-1.928	.166
12	103.000	239.000	-.441	.728
13	82.000	218.000	-1.439	.224
14	103.000	239.000	-.441	.728
15	87.000	223.000	-1.272	.313
16	95.000	231.000	-1.200	.498
17	63.000	199.000	-2.658	.043
18	62.000	198.000	-2.544	.038
Not corrected for ties.				

O.13: Mann- Whitney U Test Results for Delayed Recall-

Design 1 Original Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.
1	69.000	205.000	-2.140	.077
2	84.000	220.000	-1.394	.257
3	78.000	214.000	-1.844	.166
4	77.000	213.000	-1.781	.154
5	80.000	216.000	-2.258	.193
6	70.000	206.000	-2.198	.085
7	103.000	239.000	-.720	.728
Not corrected for ties.				

O.14: Mann- Whitney U Test Results for Delayed Recall-

Design 2 Original Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	105.000	241.000	-.337	.790
2	105.000	241.000	-.337	.790
3	99.000	235.000	-.647	.608
4	100.000	236.000	-.628	.637
5	103.000	239.000	-.720	.728
6	101.000	237.000	-.624	.667
7	110.000	246.000	-.141	.951
Not corrected for ties.				

O.15: Mann- Whitney U Test Results for Delayed Recall-

Design 3 Original Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. [2*(1-tailed Sig.)]
1	65.000	201.000	-2.340	.052
2	69.000	205.000	-2.140	.077
3	65.000	201.000	-2.340	.052
4	101.000	237.000	-.624	.667
5	63.000	199.000	-2.658	.043
6	71.000	207.000	-2.325	.093
7	79.000	215.000	-1.979	.179
8	63.000	199.000	-2.658	.043
9	104.000	240.000	-1.069	.759
Not corrected for ties.				

O.16: Mann- Whitney U Test Results for Delayed Recall-

Design 4 Original Scoring System

Item	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (1-tailed Sig.)
1	60.000	196.000	-2.500	.031
2	68.000	204.000	-2.130	.070
3	32.000	168.000	-4.071	.001
4	46.000	182.000	-3.232	.005
5	88.000	224.000	-1.919	.334
6	80.000	216.000	-2.258	.193
7	39.000	175.000	-3.634	.193
8	95.000	231.000	-1.200	.498
9	55.000	191.000	-2.984	.017
10	39.000	175.000	-3.634	.002
11	75.000	211.000	-1.779	.131
12	77.000	213.000	-1.781	.154
13	68.000	204.000	-2.130	.070
14	77.000	213.000	-1.781	.154
15	53.000	189.000	-2.856	.013
16	96.000	232.000	-1.539	.525
17	87.000	223.000	-1.609	.313
18	63.000	199.000	-2.658	.043
Not corrected for ties.				