1-1-2014

Relationships between Tests of Visual Memory in Patients with Mild Cognitive Impairment and Alzheimer's Disease

Guy Bernard deBros

Relationships between Tests of Visual Memory in Patients with Mild Cognitive Impairment and Alzheimer's Disease

by

Guy B. deBros

Presented to the Faculty of the

Graduate Department of Clinical Psychology

George Fox University

in partial fulfillment

of the requirements for the degree of

Doctor of Psychology

in Clinical Psychology

Newberg, Oregon

September, 2014

Relationships between Tests of Visual Memory in Patients with

Mild Cognitive Impairment and Alzheimer's Disease

by

Guy B. deBros

has been approved

at the

Graduate Department of Clinical Psychology

George Fox University

as a dissertation

for the PsyD degree

Signatures:

Wayne V. Adams, PhD, Chair

Members:

Kathleen Gathercoal, PhD

Marie-Christine Rutter-Goodworth, PhD

9/15/2014

Date

Relationships between Tests of Visual Memory in Patients with Mild Cognitive Impairment and Alzheimer's Disease

Guy B. deBros

Graduate Department of Clinical Psychology

George Fox University

Newberg, Oregon

September, 2014

Abstract

Alzheimer's disease (AD) is a degenerative neurological disorder characterized by cognitive and functional impairment (Budson & Solomon, 2011). Its prevalence is expected to rise in the upcoming decades as the world's population ages (Alzheimer's Association, 2014). Amnestic mild cognitive impairment (MCI) is a clinical diagnostic entity that may represent very early AD (Morris et al., 2001). Both disorders involve significant impairment in episodic memory, necessitating reliable memory measures when making diagnoses. Although verbal memory is most often impaired in the earliest stages of disease (Budson & Price, 2005), visual memory is also predictor of AD (Kawas et al., 2003). Partly due to a lack of comparative data between visual memory measures, they are often chosen based on clinical rather than psychometric needs. This study compared several commonly-used visual memory measures in order to provide data to which clinicians can refer when choosing between measures when time is limited.

A 2 × 8 mixed within-between groups design was used to compare measures in a group of 20 older adult patients at a memory disorders clinic diagnosed with AD or MCI and in a comparison group 20 normal healthy controls. Subtests from the Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2; Sheslow & Adams, 2003), the Wechsler Memory Scale, 4th Edition (WMS-IV; Wechsler, 2009), and the Brief Visuospatial Memory Test, Revised Edition (BVMT-R; Benedict, 1997) were administered to each participant as part of a comprehensive test battery. Correlations between measures were stronger in the control group (r =-.335 to .871, mean r = .157) than in the control group (r = -.421 to .825, mean r = .289). Mean scaled scores differed significantly between groups on most measures, with large effect sizes (d = 1.26 to 2.77, mean d = 1.815). Differences in mean scaled scores between some measures were larger in the clinical group than in the control group. Results demonstrate convergent validity between measures despite differences in item content and response format. Certain measures demonstrated psychometric characteristics that may be advantageous depending on the clinical setting.

Table of Contents

Approval Page ii

Abstract iii

List of Tables viii

List of Figures xi

Acknowledgements xii

Chapter 1: Introduction 1

Memory 1

Memory and Aging 2

Alzheimer's Disease 3

Pathophysiology of AD: 5

Diagnosis of AD 6

AAMI, Mild Cognitive Impairment, and the AD Continuum 8

Memory Testing: Verbal and Visual 8

Comparing and Choosing Measures 12

Existing Comparisons between Measures 14

Do Visual Memory Tests Measure Visual Memory? 17

Purposes of the Present Study 21

Rationale 22

Hypotheses 24

Chapter 2: Method 26

Participants 26

Inclusion/exclusion criteria and rationale for combined MCI/AD sample 27
Demographic characteristics 29
Clinical group 30
Control group 31
Comparisons between groups 31
Measures 33
Wide Range Assessment of Memory and Learning,
2nd Edition (WRAML2) 33
Wechsler Memory Scale, 4th Edition (WMS-IV) 34
Brief Visuospatial Memory Test, Revised Edition (BVMT-R) 36
Procedure 41
Chapter 3: Results 45
Comparison of Groups' Performance to Each Other and to Normative Data 45
Comparison of Clinical Group Performance to similar Clinical Samples 48
Hypotheses 1 and 2: Relatedness Among Visual Memory Measures 54
Recognition measures 56
Hypotheses 3 and 4: Comparisons of Mean Scaled Scores between Measures 58
Recognition measures 71
Chapter 4: Discussion 76
Conclusions/Implications 81
Other Useful Findings 84
Limitations 85

References 90

Appendix A: DSM-IV-TR Diagnostic Criteria for
Dementia of the Alzheimer's Type (DAT) 106

Appendix B: NIA-AA Diagnostic Guidelines for Probable Alzheimer's Disease 108

Appendix C: DSM-5 Diagnostic Criteria for Major and
Mild Neurocognitive Disorders 111

Appendix D: Informed Consent to Participate in Clinical Research 115

Appendix E: Curriculum Vitae 117

List of Tables

Table 1: Comparison of Performance on BVMT-R in Several Clinical Studies.....18

Table 2: Demographic Characteristics by Group.....30

Table 3: Correlations Between Demographic Variables by Group 32

Table 4: Scaled Score Differences between AD Clinical Sample and Matched Controls on Selected WRAML2 Subtests 35

Table 5: Scaled Score Differences between AD Clinical Sample and Matched Controls on Selected WMS-IV Subtests.....37

Table 6: Correlations between BVMT-R Scores and Scores on Similar Visual Memory Measures.....39

Table 7: Mean Differences Between AD and Mixed Dementia Clinical Samples and Normal Controls for Selected BVMT-R Scores 40

Table 8: Comparison of Time Points, Item Content, and Retrieval Paradigm between Visual Memory Measures Used in the Present Study 41

Table 9: Mean Scaled Score Differences Between Clinical Group and Control Group.....47

Table 10: Mean Differences between Control Group and Population Mean of 10 on Selected Visual Memory Measures 48

Table 11: Mean Differences between Clinical Group and Population Mean of 10 on Selected Visual Memory Measures 49

Table 12: Comparison of Correlation Coefficients between Control Group and Normative Data Reported by Test Authors.....50

Table 13: Comparison of Correlation Coefficients between Clinical Group and Normative Data Reported by Test Authors....51

Table 14: Clinical Group Compared to AD Clinical Group reported in the WRAML2 Test Manual....52

Table 15: Clinical Group Compared to MCI and AD Clinical Groups reported in the WMS-IV Test Manual....53

Table 16: Clinical Group Compared to AD Clinical Group reported in the BVMT-R Test Manual....54

Table 17: Intercorrelations of Scaled Score Performance for the Control and Clinical Groups....55

Table 18: Correlations between Visual Memory Mean Scaled Scores and Demographic Variables by Group....56

Table 19: Partial Intercorrelations of Scaled Score Performance for the Control and Clinical Groups....57

Table 20: Comparison of Correlation Coefficients between Clinical Group and Control Group....60

Table 21: Pair-wise Comparisons between Measures in the Overall Sample Controlling for Age, Education, and Estimated Premorbid IQ....66

Table 22: Pairwise Comparisons between Measures in the Control Group....68

Table 23: Pairwise Comparisons between Measures in the Clinical Group....70

Table 24: Mean Scaled Score Differences between AD and MCI Patients in the Clinical Group....72

Table 25: Sensitivity and Specificity of WMS-IV

Visual Reproduction Recognition Raw Scores.....73

Table 26: Sensitivity and Specificity of BVMT-R Recognition False Alarms Raw Scores 74

List of Figures

Figure 1: Mean Scaled Scores on Visual Memory Measures by Group.....64

Acknowledgements

The English word *dissertation* comes from the Latin for *path.* I was once told that "the path is wide," referring to the fact that there are many ways to reach the same destination. I took the statement rather too literally, and I meandered up, down, and across the path until the destination appeared so far out of reach that only divine intervention could point me in the right direction. Perhaps I will someday reach the destination. In the meantime, I am satisfied knowing that there is plenty of path left to explore.

The stack of paper you hold in your hands, or the electronic document projected in front of your eyes, is the product of the ceaseless encouragement of a dedicated team of mentors. In particular, I thank Wayne Adams, Kathleen Gathercoal, and Marie-Christine Rutter-Goodworth, for their contributions to my academic development; Cynthia Murphy, Diana Michalczuk, Todd Solomon, Catherine Caplis, Haley Amoia-Post, and the staff of The Memory Clinic in Bennington, Vermont, for their financial and in-kind contributions to the endeavor; Fred and Jane deBros, and Bill and Mary O'Neil, for their unconditional love; Kathryn deBros, along with Karma, Sophie, and Sandy, for walking with me along the treacherous road that led to where I am today; and God, who watches over us all with the knowledge that we belong to Him.

Having given thanks, I also dedicate the product and the process of my research to the courageous patients and caregivers who were willing to join me in my quest for knowledge. While I cannot fully understand their suffering or their joy, I hope that I have contributed in some small way to the improvement of their lives and the lives of those who suffer and are joyful with them.

Chapter 1

Introduction

Memory

A quote attributed to Endel Tulving, one of the best-known researchers in the field of memory, states, "Remembering, for the rememberer, is mental time travel." Memory is at once fundamental to the conscious experience of existence as well as enormously complex, underpinning almost all of the most important aspects of what it means to be human. It is, in a phrase, the process by which we store the information that allows us to create meaning in the world. Because of its complex interrelationships with so many cognitive processes, memory represents one of the most challenging pursuits in psychological research. Accordingly, it has been a subject of study not only from the time when psychology was a nascent science, but also from the dawn of history (Jaynes, 1976).

Ebbinghaus (1913) is memorialized as the first scientist to systematically evaluate memory processes in a human subject—himself. He created the idea of a *forgetting curve*, and provided quantification of the rate of forgetting with the passage of time. Karl Lashley induced brain lesions in laboratory animals in search of the *engram*, the name he gave to the hypothetical unit of memory storage; he concluded, having failed to find it, that memories are stored diffusely throughout the cerebral cortex (Eichenbaum & Cohen, 2001). In the 1960s, Atkinson and Shiffrin (1968) developed what they called the *multi-store model* of memory that is comprised of *sensory*, *short-term*, and *long-term* memory. Shortly after the publication of Atkinson and Shiffrin's model, Tulving (1972) clarified the difference between *episodic* and *semantic*

memory. Taken together, these theories have been extraordinarily influential in our understanding of the hippocampus, the subcortical structure whose primary function is to encode sensory information into short-term memory stores and then consolidate it into long term memory to be stored elsewhere in the brain (Milner, Squire, & Kandel, 1998).

Changes in memory occur as part of the normal aging process (Cullum, Butters, Tröster, & Salmon, 1990). Episodic memory, in particular, is susceptible to impairment when age, injury, or illness disturbs the normal functioning of the human brain.

Memory and Aging

The ancient Greeks, and Socrates in particular, were aware of the ways in which memory may decline in old age (Berchtold & Cotman, 1998). They, and most of their intellectual descendents, misattributed the proximal cause of memory decline to advancing age. Hardly anyone born before 1900 could expect to live much beyond the age of 65. It was therefore difficult for ancient and pre-modern scientists to compare normal and abnormal memory functioning in the oldest old because the base rate of survival was so low. Until well into the 20th century, most people believed that cognitive decline was an inevitable side effect of the aging process. We now know this to be false.

A decline in memory does not necessarily indicate the presence of pathology. V. A. Kral (1962) was the first researcher to classify *senescent forgetfulness* as either *benign* or *malignant*. While many people experience a change in cognition as they age, the change is only sometimes due to a disease process; age itself is not a disease. Kral hypothesized that age-related cognitive changes may be due to a mild degree of general cortical atrophy that progresses slowly over time, in the absence of the severity of hippocampal atrophy that is seen in degenerative

neurological disorders. Kral's benign senescent forgetfulness is now more frequently called age-associated memory impairment (AAMI). In support of Kral's theoretical stance, Walhovd and colleagues (2005) demonstrated that age-related volumetric changes in the hippocampus are curvilinear over the lifespan, first growing and then shrinking with increasing age.

Malignant senescent forgetfulness, in the presence of certain diagnostic criteria (see Appendices A and B), is now called *major neurocognitive disorder* or *dementia*. Dementia is a general term that implies a deviation from the normal aging process (Grundman et al., 2004). In the lay public, dementia is used synonymously with Alzheimer's disease (AD). While AD is the most common cause of dementia in people over the age of 65 (Budson & Solomon, 2011), it is only one of the dozens of known causes of dementia.

Alzheimer's Disease

Alois Alzheimer's description of Auguste Deter, *Über eine eigenartige Erkrankung der Hirnrinde* ("Regarding a peculiar illness of the cortex," 1907; Stelzmann, Schnitzlein, & Murtagh, 1995), has been immortalized in the literature as the index case of his eponymous disease (Berchtold & Cotman, 1998). Using a microscope, the most sophisticated method of evaluation at the time, Alzheimer documented post-mortem histopathological evidence of amyloid plaques and neurofibrillary tangles in the brain of a woman who died in her early 50s. Only a century later did Müller and colleagues (Müller, Winter, & Graeber, 2013) demonstrate that hers was a case of a very early-onset, autosomal-dominant form of AD that is extraordinarily rare in the general population. Nevertheless, Alzheimer's findings became the definitive metric by which the idiopathic form of AD, the kind most commonly seen in older adults with progressive memory decline, is diagnosed.

AD is a degenerative disease that represents a departure from typical aging. By current estimates, more than 5 million Americans currently have AD (Alzheimer's Association, 2014); of those 5 million, up to 50%, or 2.5 million, have not been diagnosed. The Alzheimer's Association estimates that by 2050 there may be as many as 1 million new cases of AD per year, adding up to a prevalence of almost 14 million in the United States alone. Of the leading causes of death in the United States, AD is the 6th overall, and the 5th in people over the age of 65. Concurrent with a decrease in the proportion of deaths resulting from heart disease, stroke, and prostate cancer in the decade ending in 2010, there was a 68% increase in the proportion of deaths resulting from AD. The estimated cost of treating people over the age of 65 with Alzheimer's disease is expected to rise to $214 billion per year in 2014, an increase of 5% from the previous year; this does not include the estimated contributions of more than 15 million unpaid caregivers, whose time and efforts were valued at $220 billion in 2013 (Alzheimer's Association, 2014).

Clearly, AD is an important public health issue in a country with almost 40 million inhabitants who are older than 65 (American Psychological Association, 2012). In the scientific community, research has focused on the increasingly early detection of AD with the hope that intervening at a very early stage in the disease process may prevent some of the symptoms that eventually develop as AD progresses. Because memory is one of the first cognitive domains in which decline is evident in AD patients, the early detection of memory impairment is a crucial step in identifying cases of AD in order to manage the societal—and especially the personal—damage that AD can cause.

Pathophysiology of AD: The Amyloid Cascade Hypothesis, Amyloid-β, and *Tau*

The amyloid precursor protein (APP), which is present in the membranes of neurons, is normally cleaved by alpha- and gamma-secretase, and the byproducts are cleared by cerebrospinal fluid (CSF; Querfurth & LaFerla, 2010). When APP is cleaved by β-secretase before being cleaved by γ-secretase, however, a monomeric peptide composed of 40–42 amino acids called amyloid-β is produced. The biological role, if any, of amyloid-β is unknown, although a suggestion has been made that it may be an antimicrobial molecule (Soscia et al., 2010). For reasons that are not fully understood, amyloid-β monomers tend to join together to form dimers that gather into oligomers, which are toxic to cells and are not easily cleared by CSF. The buildup of amyloid-β oligomers leads to the formation of amyloid plaques, which also contain the detritus of dead cells. Amyloid plaques interfere with synaptic transmission between neurons. In addition, they are thought to cause the hyperphosphorylation of τ (*tau*), a protein that is present in the microtubules of neuronal axons. In its normal form, τ provides structure to the microtubules. When hyperphosphorylated, however, τ causes the microtubules to twist into paired helical filaments. Neurons die off due to impaired intracellular nutrient transport, and the axons of these dead cells containing hyperphosphorylated τ form neurofibrillary tangles.

Amyloid plaques and neurofibrillary tangles accumulate predominantly in the hippocampus, the temporal lobes, and the parietal lobes, leading to the characteristic cognitive changes associated with Alzheimer's disease: memory impairment, word-finding difficulty, spatial confusion, and disorientation (Budson & Price, 2005). The pathophysiological process associated with cognitive symptoms is believed to begin 10 to 20 years before symptoms are first noticed by patients or family members (Budson & Solomon, 2011).

Acetylcholinergic neurons are a particular target of this degenerative process. As these cells die in increasing numbers, post-synaptic acetylcholine receptors upregulate in order to compensate for the decrease in acetylcholine at the synapse. Meanwhile, acetylcholinesterase, an enzyme that is partially responsible for the removal of acetylcholine from the synapse, remains active, limiting the availability of acetylcholine at post-synaptic receptors. Acetylcholinesterase inhibitors are thought to be therapeutic due to their action at the synapse, intensifying neuronal transmission by *reversibly* (i.e., temporarily) inhibiting the breakdown of acetylcholine (Rogers, Farlow, Doody, Mohs, & Friedhoff, 1998). As neurons continue to die, compensatory mechanisms begin to fail, leading to an increasingly rapid decline in cognition and function. Alzheimer's disease is almost always fatal, usually within 6-8 years of diagnosis (Budson & Solomon, 2011), typically due to complications that arise from functional impairment, such as bedsores, pneumonia, aspiration, or falls.

Diagnosis of AD

Several diagnostic frameworks have been used to diagnose AD; in general, they all share the same core criteria: a progressive decline in memory along with impairment in one other cognitive area and the presence of functional decline (i.e., impairment in activities of daily living). It is important to note that while all cases of AD are dementias by definition, not all cases of dementia are due to AD; only around 70% of dementias are due to AD (Budson & Solomon, 2011). In addition, all clinical diagnoses of AD should be considered diagnoses of exclusion; although the biomarkers discussed below can add certainty to clinical judgment, the final diagnosis is made by histopathological analysis on autopsy.

In 1984, the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA) published diagnostic criteria for Alzheimer's disease (McKhann et al., 1984). The resulting NINCDS-ADRDA criteria allowed for the diagnosis of possible or probable AD, with disease stage classified as early (mild), middle (moderate), and late (severe) depending on cognitive and functional impairment. The Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR; American Psychiatric Association, 2000) uses the term *Dementia of the Alzheimer's Type* (DAT) in describing the symptoms of the cognitive and functional impairment that constitute the syndrome (Appendix A). In 2011, the National Institute on Aging (NIA) and the Alzheimer's Association (AA, formerly known as the ADRDA) published a new set of guidelines in three parts (Albert et al., 2011; McKhann et al., 2011; Sperling et al., 2011) as an update to the NINCDS-ADRDA criteria. The NIA-AA criteria provide for the diagnosis of MCI due to the AD pathophysiological process. Two additional stages on the AD spectrum were added: presymptomatic/prodromal/preclinical AD and symptomatic but non-demented AD (i.e., MCI). The new guidelines also encourage the use of biomarkers, such as CSF analysis and amyloid positron emission tomography (PET), to increase the accuracy of the diagnosis of preclinical AD (Budson & Solomon, 2012). A summary of the new guidelines can be found in Appendix B. New criteria for Major and Minor Neurocognitive Disorder in the DSM-5 (American Psychiatric Association, 2013) closely resemble the NIA-AA diagnostic criteria and appear to have properly conjoined the disparate diagnostic criteria promoted by various medical specialties into one cohesive framework. These criteria are summarized in Appendix C.

AAMI, Mild Cognitive Impairment, and the AD Continuum

Larrabee, Levin, and High (1986) demonstrated evidence in support of age-associated memory impairment (AAMI) as a clinical entity that is distinct from what, at the time, was called senile dementia. In a sample of 88 normal controls between the ages of 60 and 90, they found that between 10% and 20% of the sample exhibited memory impairment beyond what would be expected for their age but in the absence of other cognitive impairment. These participants nevertheless performed better overall than a group of AD patients. While the authors conclude that their data support Kral's (1962) concept of benign senescent forgetfulness, it is important to note that the idea of mild cognitive impairment was not widely endorsed by memory researchers of that era. It is therefore difficult to determine whether the participants with so-called AAMI may have demonstrated the degree of memory impairment that would earn a diagnosis of MCI today.

There is growing consensus that mild cognitive impairment (MCI) represents a very early stage of AD (Albert et al., 2011; Morris et al., 2001), in which memory impairment is present in the absence of functional impairment. Patients with MCI and AD have similar patterns of memory impairment, and non-memory measures (e.g., ADLs) better distinguish one from the other (Jacova, Kertesz, Blair, Fisk, & Feldman, 2007). At times the distinction between the two is made on clinical rather than psychometric grounds, and the treatment for both is similar (Budson & Solomon, 2011).

Memory Testing: Verbal and Visual

Neuropsychological testing is a critical part of the diagnosis of AD (McKhann et al., 1984; McKhann et al., 2011; Smith & Bondi, 2013). The American Psychological Association,

in its *Guidelines for the Evaluation of Dementia and Age-Related Cognitive Change* (2012), declared that "Psychologists are aware that standardized psychological and neuropsychological tests are important tools in the assessment of dementia and age-related cognitive change" (p. 2). AD impairs both encoding (short-term to long-term memory storage) and retrieval (Budson, Wolk, Chong, & Waring, 2006). Memory impairment is one of the earliest sign of impending AD/dementia (Jacova et al., 2007) and "episodic memory tests appear to have the greatest predictive accuracy" (p. 311). For various reasons—perhaps in particular due to the relative stability of verbal memory with normal aging (Cullum et al., 1990; Giambra, Arenberg, Kawas, Zonderman, & Costa, 1995; Salthouse, 2010), as well as the fact that anomia is a prominent early feature of AD (Budson & Price, 2005) and the importance of verbal communication in modern society—verbal memory has been a consistent focus of study in the dementia literature. However, an evaluation that does not include at least one visual memory measure may not adequately identify all cases of amnestic MCI (Smith & Bondi, 2013); up to 70% of all MCI cases eventually convert to AD at a rate of 10-15% per year (Budson & Solomon, 2011), making visual memory an important factor in the early detection of AD pathology. Given some findings of asymmetric patterns of atrophy in AD (Derflinger et al., 2011), and because visual memory is presumed to be localized to the right hemisphere (Lezak, Howieson, & Loring, 2004), a neuropsychological evaluation of memory impairment cannot be considered complete without the inclusion of at least one measure of visual/non-verbal memory.

There is evidence to suggest that visual memory may also decline more quickly than verbal memory with increasing age (Skilbeck & Woods, 1980; Riege & Inman, 1981). Verbal memory, on the other hand, appears to be more resistant to decline with age (Burkhart, 2011;

Gale, Baxter, Connor, Herring, & Comer, 2007; Sheslow & Adams, 2003). Other studies have failed to confirm the idea of differential cognitive decline with age (Salthouse, Fristoe, & Rhee, 1996). It is also unclear to what extent verbal abilities mediate performance on visual memory measures, although it seems likely that this varies depending on the chosen measure. Bornstein and Chelune (1989), in a factor analysis of the WMS-R, found that the loading of non-verbal memory on a verbal ability factor increased with age. This would suggest that older adults may rely more heavily on verbal memory and skills (which presumably decline more slowly) than their younger counterparts.

Performance on visual memory measures has been shown to differentiate between groups of AD patients and normal controls. In a meta-analysis of 47 studies that included 1,207 preclinical AD and 9,097 controls, Bäckman, Jones, Berger, Laukka, and Small (2005) found an episodic memory effect size of about 1.03; within the domain of episodic memory, the largest effect sizes were found on delayed recall tasks. Troyer and colleagues (2008) used a modified scoring procedure to show that, in a sample of aMCI patients and normal controls, associative memory showed a larger between-groups difference than item memory.

Visual memory impairment can also predict conversion to AD in normal and memory-impaired participants. In population studies, visual memory impairment can be detected up to 10 years preceding a diagnosis of AD (Kawas et al., 2003; Tierney, Yao, Kiss, & McDowell, 2005). A population-based study conducted by Albert and colleagues (2001) measured a group of 165 participants with and without memory complaints. Participants were classified at baseline as either cognitively normal or questionable AD. At a 3-year follow-up visit, participants were then classified as either cognitively normal, questionable AD, or probable AD. All participants

completed 20 neuropsychological measures (covering the domains of memory, language, executive function, and visuospatial skills) at both time points. Using a stepwise discriminant function analysis, the authors demonstrated that performance on one measure of visual memory at baseline accurately differentiated participants who converted from normal to AD and from questionable AD to AD.

In a sample of 145 normal controls and AD patients, Swainson and colleagues (2001) found that the Paired Associates Learning task of the Cambridge Neuropsychological Test Automated Battery (CANTAB), a visual memory measure, differentiated normal controls from AD patients and depressed patients from AD patients, and classified certain questionable AD patients as actually having AD. Blackwell and colleagues (2004) demonstrated, in a sample of 43 patients with questionable AD followed over 32 months, that 11 participants (26%) converted to AD while 29 participants (67%) remained in the questionable AD group; the CANTAB Paired Associates Learning task predicted conversion to AD with 100% accuracy. A measure of immediate visual memory reliably differentiated slow and fast progression of cognitive decline (Buccione et al., 2007). Eslinger, Damasio, Benton, and Van Allen (1985) compared a sample of patients with dementia due to a variety of etiologies and found that a combined factor that included visual retention, as well as verbal fluency and orientation to time, classified participants as normal or demented with 89% accuracy.

Findings such as these have been extended to patients with MCI. In a sample of 166 consecutive patients referred to a memory clinic, Alladi and colleagues (2006) found that out of 124 patients who met inclusion criteria (not demented, not depressed), 72 of them (58%) were classified as having MCI when Petersen criteria (Petersen et al., 1999) were used (which judges

episodic memory deficits using only verbal memory measures), while 90 of them (73%) were classified as having MCI if episodic memory impairment was documented with either verbal memory measures or a visual memory measure (CANTAB Paired Associates Learning). Using the DMS48, a measure of visual recognition memory, De Anna and colleagues (2014) studied a group of 33 patients with amnestic MCI and 26 normal controls and found that impaired recognition memory at baseline predicted greater decline in cognition after 18 months; all 3 participants who converted to AD belonged to the group that was impaired at baseline. Thorough reviews of visual/nonverbal memory testing have been written by (1997) and Iachini, Iavarone, Senese, Ruotolo, and Ruggiero (2009).

Comparing and Choosing Measures

Episodic memory is not a unitary construct. Theory and practice suggest that it can be subdivided into verbal and visual/nonverbal processes. The construct of visual/nonverbal memory can be measured in different ways, which can influence whether and how it relates to verbal memory and episodic memory in general. The multi-store model of memory (Atkinson & Shiffrin, 1968) proposes a processing system that encodes, retains, and retrieves incoming perceptual information. Information is encoded, retained, and retrieved separately in conjunction with how the information is perceived. Indeed, Phillips and Christie (1977) found that visual memory has STM and LTM characteristics that are consistent with the Atkinson-Shiffrin (Atkinson & Shiffrin, 1968) model.

Immediate recall measures assess the degree to which information has been properly encoded, while delayed recall tasks measure how much information was retained and how well it was retrieved. Assuming adequate encoding, differential performance on uncued (recall) and

cued (recognition) delayed retrieval tasks may clarify whether impairment in retrieval is due to impaired retention or retrieval difficulties. Even recognition tasks can be subdivided into multiple-choice, yes-no recognition, or forced-choice (either-or) recognition to further clarify the nature of a retrieval deficit. Information to be processed may be not only verbal or nonverbal in nature but also simple or complex, concrete or abstract, meaningful or rote.

When comparing tests to one another, issues of validity emerge. In general, validity is demonstrated by evidence that the test measures what it claims to measure (Gregory, 2007). Face validity suggests that a test appears to measure what it actually does measure. For example, a visual memory test most likely involves asking the examinee to memorize some visual information and recall it later. A list-learning task in which words are read aloud to the examinee is not face valid as a measure of visual memory. Another form of validity, one that is more easily demonstrated in research, is construct validity, or evidence in support of group differences that are consistent with an underlying theory of the construct being measured. If visual memory is impaired in patients with right temporal lesions, and if those patients scores on visual memory measures differ significantly from the population mean, construct validity has been demonstrated. When multiple tests are available, each of which has adequate construct validity, they may be compared to each other to demonstrate convergent validity, or evidence that test scores correlate with each other when administered to the same group of people, indicating that they all measure a similar construct in similar ways. Finally, there is ecological validity, a form of validity that is becoming increasingly important. A test may be considered ecologically valid if performance on the test in a clinical setting (strength or weakness) is related to real-world functioning. For example, a visual memory measure may be considered ecologically valid if

people who perform poorly on it also perform poorly on an informant-reported measure such as ADLs. In terms of ecological validity, visual memory measures are relevant to some of the problems that people with MCI and AD most frequently face, including misplacing objects around the house or getting lost while driving.

Visuospatial tasks can also differentiate AD from Parkinson's disease dementia or dementia with Lewy bodies (two diseases whose underlying pathophysiology is presumed to differ significantly from AD) and frontotemporal dementia (again, a disease with a different underlying etiology) from AD (Jacova et al., 2007; Pachana, Boone, Miller, Cummings, & Berman, 1996) making them a useful addition to a neuropsychological test battery when underlying etiology is uncertain.

Existing Comparisons between Measures

Several past studies have compared visual memory measures, although none directly compared the subtests and measures proposed for this study. A doctoral dissertation completed by Hall (2006) compared the WRAML2 to the WMS-III in a clinical sample of diagnosed AD patients and found that, with the exception of verbal memory and overall memory index scores, the WRAML2 and the WMS-III were generally *not* equivalent. However, Hall demonstrated differences between visual memory *indexes* (WRAML2 Visual Memory vs. WMS-III Visual Immediate, $d = .58$, and WRAML2 Visual Recognition vs. WMS-III Visual Delayed, $d = .72$). Using a Bonferroni correction for multiple comparisons, the difference between WRAML2 Visual Recognition and WMS-III Visual Delayed remained significant. Correlations were also reported: WRAML2 Visual Memory and WMS-III Visual Immediate, $r = .64$; WRAML2 Visual

Memory and WMS-III Visual Delayed, $r = .21$; WRAML2 Visual Recognition and WMS-III Visual Immediate, $r = .44$; WRAML2 Visual Recognition and WMS-III Visual Delayed, $r = .77$.

There are several important differences between Hall's study and the present study. Her sample was actively recruited from the community rather than passively recruited in a clinical setting; the sample included a disproportionate number of female participants, despite evidence that the incidence of AD is approximately equal for males and females (Alzheimer's Association, 2014); participants were already diagnosed with Alzheimer's disease in the mild to moderate stages (no MCI patients were included); some were being treated with medications, but it was not clear which medications or at which doses; participants were included only if they scored between 18 and 23 on the MMSE, potentially omitting those who scored between 24 and 26 who may have had memory and functional impairment consistent with a diagnosis of AD; the WMS-III rather than the WMS-IV was used because the WMS-IV had not been published at the time; and participants with depression were excluded, despite evidence that depression can be a symptom of early AD rather than a confounding causal factor.

Frise (2009) found that a difficult visual memory task (the Rey-Osterrieth Complex Figure Test; RCFT; Rey, 1941), as typically scored, had too high of a floor to measure meaningful differences in older adults. Burkhart (2011) extended this finding to AD patients. Neither study made comparisons between multiple visual memory measures, but both suggest that as impairment worsens, the importance of variability in scores at the lowest levels of performance increases, as measures that generate a broader range of age-controlled scaled scores are more useful in making clinical decisions.

In a study of 113 patients, Golden and colleagues (2005) compared differences between groups of patients with vascular dementia and AD and found that the vascular dementia group performed better overall, with "similar scores on complex tests and different scores on basic tests" (p. 1570). Correlations between measures were not reported, and no normal controls were studied. Cherner and colleagues (2007) collected normative data for the BVMT-R in a sample of 127 Spanish-speaking participants of Mexican descent and noted the importance of including education as a demographic correction when using the BVMT-R in this population. Their sample ranged in age from 18 to 79 years; the mean age was 38.6 years (SD = 18.0), and they cautioned that their data should not be used with older Spanish-speaking adults.

Gale, Baxter, Connor, and Herring (2007) administered Form 4 of the BVMT-R to a sample of 172 cognitively normal adults between the ages of 60 and 89. Normative data were provided for 14 overlapping age groups, reflecting the structure of normative data provided by the BVMT-R professional manual. In an unpublished doctoral dissertation, Gurczynski (2009) administered Form 1 of the BVMT-R to a sample of 49 cognitively normal adults above the age of 80. Normative data were provided for two age groups (80–84: $N = 36$; 85–89: $N = 13$). In a doctoral dissertation that was subsequently published in a peer-reviewed journal, Kane (2012; Kane & Yochim, 2014) administered Form 1 of the BVMT-R to a sample of 175 participants (109 normal controls, 49 residents of independent or assisted living communities, 17 patients at community mental health center). Scaled scores for Total Recall were significantly correlated with age, $r = -.29$, $p < .01$, and education, $r = .35$, $p < .01$). Scaled scores for Delayed Recall were also correlated with age, $r = -.32$, $p < .01$, and education, $r = .38$, $p < .01$. Normative data were provided for a subset of participants in either the normal group or the assisted-living group

who were aged 80 or older and were judged to be cognitively normal ($N = 59$), grouped by age and level of education. However, data for all education levels were only provided for the 80–88-year-old age group ($N = 29$).

An unpublished doctoral dissertation by McCoy (2004) evaluated intra-individual variability in cognitive scores over 15 days in participants with amnestic MCI ($N = 15$) and normal controls ($N = 53$) over the age of 65. Participants were administered the BVMT-R as part of a comprehensive neuropsychological test battery on the first visit only. Means and standard deviations were reported for BVMT-R Immediate raw scores and BVMT-R Delayed *T* scores in each group. The mean BVMT-R Immediate raw score was approximately 9 points lower (out of a possible 36 points) in the MCI group than in the control group, $t(66) = -5.113$, $p < .001$, $d = -1.53$. The mean BVMT-R Delayed *T* score was 16.7 points lower in the MCI group than in the control group, $t(66) = 6.546$, $p < .001$, $d = -1.88$. No other visual memory measures were compared.

Means and standard deviations for scores in these studies relevant to the present study are presented in Table 1. With the exception of these studies, few data are available comparing these measures to one another.

Do Visual Memory Tests Measure Visual Memory?

The literature presents conflicting views of whether visual memory tests actually measure visual and not verbal or general memory. In other words, divergent validity between verbal and visual memory tests is sometimes difficult to demonstrate. Perhaps because of its status as one of the most widely used memory measures (Strauss, Sherman, & Spreen, 2006), the Wechsler Memory Scale appears to be the measure that is most frequently subjected to factor analytic

Table 1

Comparison of Performance on BVMT-R in Several Clinical Studies

			BVMT-R Score			
			BVMT-R Immediate		BVMT-R Delayed	
Study	Age Range	*N*	Mean	*SD*	Mean	*SD*
Gale et al. (2007)	80–84	41	14.8	6.0	6.5	2.8
	82–86	29	14.6	6.0	6.3	2.7
	84–88	16	13.1	5.0	5.4	2.7
	86–89	19	14.2	5.2	6.3	2.7
Gurczynski (2009)	80–84	36	14.72	5.90	6.03	2.47
	85–89	13	12.92	5.27	5.92	3.25
Kane & Yochim (2014)	80–88	29	15.52	5.40	6.41	2.01

Note. BVMT-R = Brief Visuospatial Memory Test, Revised Edition.

scrutiny. Several studies conducted by Leonberger, Nicks, and colleagues have failed to demonstrate the existence of a visual memory factor that is separate from a general visuospatial factor (Leonberger, Nicks, Goldfader, & Munz, 1991; Leonberger, Nicks, Larrabee, & Goldfader, 1992; Nicks, Leonberger, Munz, & Goldfader, 1992). The same pattern was demonstrated by Burton, Mittenberg, and Burton (1993).

In factor-analytic studies of verbal and visual memory, visual memory consistently loads on a separate factor (Bowden, Carstairs, & Shores, 1999; Hoelzle, Nelson, & Smith, 2011; Holdnack, Zhou, Larrabee, Millis, & Salthouse, 2011; Millis, Malina, Bowers, & Ricker, 1999;

Price, Tulsky, Millis, & Weiss, 2002; Sewell, Downey, & Sinnett, 1988; Tulsky & Price, 2003). On the other hand, Leonberger and colleagues (1992) found (on the WMS-R) that "separate verbal and visual memory components failed to emerge. Moreover, several tests intended to measure visual memory did not load on the general memory factor, loading instead with nonverbal and spatial cognitive skills".

Larrabee, Kane, and Schuck (1983) conducted a factor analysis comparing WMS and WAIS subtests in a sample of 256 normal and non-normal subjects and were not able to demonstrate evidence that the Visual Reproduction subtests load on a separate visual memory factor. The same group later demonstrated that Visual Reproduction II did load on a visual memory factor, while Visual Reproduction I loaded on a general visuospatial factor (Larrabee, Kane, Schuck, & Francis, 1985) or visual/performance IQ factor (Larrabee & Curtis, 1995). Using another visual memory measure, the Continuous Visual Memory Test (CVMT; source), the group demonstrated additional evidence for a combined immediate visual memory/nonverbal factor, while delayed visual memory emerged as a "pure" visual memory factor (Larrabee, Trahan, & Curtiss, 1992).

Bornstein and Chelune (1988), in a sample of 434 normal controls, found that verbal and visual memory loaded on separate factors only when delayed recall was included. The third factor was related to attention and IQ; a similar factor structure (verbal memory, visual memory, attention/concentration) was found in the WRAML2 normative study (Sheslow & Adams, 2003).

It is also unclear whether and how neurological diseases such as AD affect the relationship between visuospatial abilities and visual memory. Deluca and Cicerone (1991) found a combined visual ability/memory factor in a sample of 59 brain-injured patients. In a

study of 308 normal controls, 35 AD patients, and 35 Huntington's disease patients (who presumably have intact medial temporal lobes), Delis, Jacobson, Bondi, Hamilton, and Salmon (2003) found that CVLT immediate total and long delay free recall were significantly correlated in the control group (r = .81) and in the Huntington's disease group (r = .85) but only moderately in the AD group (r = .36). To address the issue of restricted range, they conducted a *post-hoc* analysis between the AD group another sample of Huntington's disease patients that was chosen to match the AD group in terms of variability in scores. California Verbal Learning Test (CVLT) immediate total and long delay free recall were still significantly correlated in the separate Huntington's disease group (r = .66). They also conducted a factor analysis of CVLT scores in a group of AD patients and found that immediate and delayed memory loaded on separate factors, whereas immediate and delayed memory tend to load on a single factor in normative studies.

In a population study, Heilbronner, Buck, and Adams (1989) were not able to separate a pure nonverbal *memory* factor from a general nonverbal factor, suggesting that visuospatial deficits may explain more of the variability in visual memory scores than visual memory itself. Of course, visual memory is strongly influenced by visuospatial skills (Heilbronner, 1992). Because AD affects the parietal lobes (Querfurth & LaFerla, 2010), which are presumed to mediate most visuospatial skills (Lezak et al., 2004), AD may lead to differentially impaired performance on visual memory measures as compared to verbal memory measures. Regarding the ease with which supposedly non-verbal content can be verbalized, Eadie and Shum (1995) demonstrated that Chinese characters are more difficult to verbalize, at least for English-speaking participants, than geometric designs. This is unlikely to be true for readers of Chinese, as Adams and deBros (2010) demonstrated evidence that there are some small differences in

visual memory performance between English-speaking and Chinese-speaking children that diminished with age.

Purposes of the Present Study

There is potential utility in alternate methods of assessing visual memory. For example, different tests have different psychometric properties that may indicate the use of one over another in certain clinical groups; tests with reliability values of less than .80 are generally unfit for clinical use, and reliabilities of .90 are expected when conducting "high stakes" testing (e.g., forensic evaluation, intellectual assessment). There are also pragmatic concerns, in that some tests are easier or faster than others to administer. Before using any new measure, however, the clinician must ascertain whether a test is equivalent to other tests that have been proven over multiple studies to be reliable and valid. This can be done informally by comparing normative data provided by test authors, but judgments made using such information is limited by the test authors' choices of clinical samples as well as the fact that different measures are compared at different times in different samples that may not be comparable. Demographic data change over time, limiting the generalizability of the results of normative studies (Flynn, 1987). Memory tests are updated on varying schedules; because of changes in administration and scoring procedures, it is necessary to demonstrate the relationships between the most recent editions of each test in order to compare them before deciding which to use for which purposes. Given so many potentially distinct memory processes and systems and even more methods by which to assess their strengths and weaknesses, practicing clinicians are left to decide on their own how to go about choosing "the best tool for the job." For clinical purposes, then, equality is best judged by a head-to-head comparison of multiple measures in the same clinical sample.

Rationale

The present study was not intended to address construct validity from a factor-analytic standpoint, but rather to clarify some important questions in order to aid clinical decision-making. In a review of visual memory measures, Moye (1997) concluded that future studies should address several key issues:

> Research should incorporate new and potentially superior design memory tests … Construct validity research should incorporate item analyses to study aspects of test stimuli that may relate to test specificity and overall reliability … Construct validity research could examine the relationship of performance on different nonverbal memory tasks with one another, with areas of brain function, and in predicting everyday function … Studies that collect and compare multiple paradigms and multiple tests for each paradigm may facilitate our understanding of the more general content domain for the construct of nonverbal memory, and the extent to which various paradigms estimate this domain (p. 167).

The present study aimed to address items 1 and 3 from Moye's (1997) list.

When working with AD and MCI patients, clinicians must measure multiple cognitive domains in a relatively short amount of time, ideally in one test session, with patients who are easily fatigued and are likely to be cognitively impaired. Because of the importance of early detection (Alzheimer's Association, 2014), the most sensitive measures of cognitive decline must be determined while awaiting additional evidence that biomarkers, such as PET using amyloid-binding radioactive markers and cerebrospinal fluid (CSF) amyloid tests, become as reliable as cognitive tests in detecting the disease (Budson & Solomon, 2011).

The diagnostic test battery used at The Memory Clinic consists of selected subtests that were chosen for their sensitivity and specificity and have remained relatively unchanged (aside from the addition of new versions of the same subtests) for the past 25 years. One of the measures used in this study (WMS-IV Visual Reproduction; Wechsler, 2009) was part of that battery until approximately two years ago, when it was arbitrarily replaced with the Brief Visuospatial Memory Test, Revised Edition (BVMT-R; Benedict, 1997). Although the Visual Reproduction subtest has decent psychometric properties, it requires a significant amount of time and effort to administer and score, and observations suggest that patients find it difficult and unpleasant to complete; scores derived from the measure, at least in the population of patients seen at the clinic, tend to fall into a bimodal distribution, limiting conclusions that can be made to binary yes-no decisions about impairment. The decision to replace it with the BVMT-R was made on practical grounds: it is particularly easy to administer and score; it is brief; it provides information about learning over multiple trials as well as delayed recall and recognition tasks; and it is less daunting to patients, at least according to the observations of examiners.

If the measures are determined to be equivalent in terms of psychometric properties, the choice can be made on pragmatic rather than statistical grounds. Otherwise, if significant differences were found between measures, empirical decisions could be made in order to maximize the effectiveness of the overall test battery. Because there are many more visual memory measures available for purchase, we decided to compare the WMS-IV and the BVMT-R to two additional measures from the Wide Range Assessment of Memory and Learning, Second Edition (WRAML2), Design Memory and Picture Memory. Each of these measures was demonstrated in normative studies to be clinically useful in differentiating AD patients from

normal controls (Benedict, 1997; Sheslow & Adams, 2003; Wechsler, 2009). It was therefore judged to be acceptable to choose one or another of these measures for use in the battery.

When the WMS-IV Visual Reproduction subtest was replaced with the BVMT-R in The Memory Clinic's test battery, little empirical evidence was available comparing the two measures. A head-to-head comparison using the same sample of patients was judged to be the best way to gather data allowing the choice of one over another. The present study was conducted to address some of the limitations of previous comparative studies and to provide additional data that did not yet exist for these particular visual memory measures. A sample of new patients referred to The Memory Clinic due to subjective memory complaints were evaluated using all of the above-mentioned visual memory measures; those who went on to receive a diagnosis of AD or MCI due to AD pathology based on the results of the evaluation were included in the clinical group. The measures were also administered to a convenience sample of normal controls in order to compare the measures to each other in a non-impaired group and to compare performance on measures between groups. The latest editions of each measure were used in order to provide results that may be useful to the practicing clinician.

Hypotheses

1. Based on normative data (Benedict, 1997; Sheslow & Adams, 2003; Wechsler, 2009), it was hypothesized that measures would be moderately to strongly correlated with each other in this sample. Theoretically, if visual memory is a unitary construct, different methods of measuring the construct should yield similar results in the same group of patients.

2. It was hypothesized that the strength of the relationships between visual memory measures within each subgroup would not differ significantly between subgroups.
3. It was hypothesized that the mean scores for each visual memory measure would differ between subgroups. As originally stated, this meant that each mean score for each individual visual memory measure would differ between subgroups.
4. It was hypothesized that differences between mean scaled scores for each measure would differ between groups. In other words, measures may be equivalent in the overall sample but not in either subgroup, or else in one subgroup but not in the other. Specifically, it was hypothesized that, while the control group would perform relatively well on most measures, larger differences in performance between measures would be found in the clinical group.

Chapter 2

Method

Participants

The present study used a cross-sectional design to compare two groups of participants who were presumably *normal* and *memory impaired*. Participants were selected from the population of clinician-referred patients at a memory disorders clinic in Bennington, Vermont. All new patients who presented for evaluation were offered an opportunity to participate in the study. The control group was selected from a population of caregivers of current patients and community-dwelling older adult volunteers who did not have subjective memory complaints when they were approached to solicit participation. The initial proposal called for caregivers of new patients (typically a spouse, child, or sibling) to participate as controls. Upon further consideration, however, some additional concerns emerged regarding that strategy. In particular, concerns were voiced by the staff of the Memory Clinic that this may place an undue burden on the families of patients who were in the process of undergoing evaluation and who would most likely receive a diagnosis of intractable neurological decline. A decision was made to approach the caregivers and relatives of patients who were already enrolled in clinical trials and who visited the clinic regularly. These caregivers were known to be interested in contributing to scientific knowledge, and the already-established relationship with them made it easier to request assistance for the present study. In addition, the parents of some staff members, most of whom were not caregivers of AD patients but who were in the desired age range and did not have subjective memory complaints, were also recruited for the control group.

Inclusion/exclusion criteria and rationale for combined MCI/AD sample. The AD research community is moving toward a consensus that amnestic MCI represents an intermediate or preclinical stage of AD pathophysiology (Albert et al., 2011; Morris et al., 2001). The underlying biological changes that precede AD begin long before cognitive symptoms are first observed (Budson & Solomon, 2011). MCI is a clinical diagnostic entity that represents a point on a continuum of cognitive changes at which impairment is present but not severe enough to produce functional deficits (Petersen et al., 1999). Patients who present for memory evaluation may lie anywhere along the continuum; the stage of pathology is not typically known at the time of evaluation. Therefore, analyses conducted on samples that include both MCI and AD patients should provide data that are useful to clinicians who may expect to see patients at varying stages of pathology. Accordingly, participants were included in the clinical sample if their initial evaluation resulted a diagnosis of either possible/probable Alzheimer's disease or MCI single- or multiple-domain amnestic type (i.e., MCI due to Alzheimer's disease pathology), in the absence of other forms of dementia. The risk of developing Alzheimer's disease increases with age, from 2.5% at age 65 to almost 50% at age 85 (Budson & Solomon, 2011), although most cases of AD are diagnosed after the age of 65 (Alzheimer's Association, 2014). On the other hand, the risk of developing MCI is not as closely tied to age. Therefore, participants of any age were screened for inclusion in the study.

Patients who present with subjective memory complaints have not usually been diagnosed and are almost never receiving treatment for AD. However, some new patients request an evaluation to confirm or clarify a diagnosis made by a primary care physician or a neurologist who may not have access to neuropsychological test results. In such cases, the patient may have

been prescribed symptomatic medication such as an acetylcholinesterase inhibitor by the referring physician. Symptomatic medications are known to have a positive effect on cognition in general (e.g., MMSE scores; Rogers et al., 1998). Therefore, participants who were already taking one or more medications approved by the FDA for the treatment of Alzheimer's disease (including off-label) at the time of their initial evaluation were excluded from the sample. Participants were also excluded if they reported taking over-the-counter cholinergic medication, such as Huperzine A (Wang et al., 2009).

There are many FDA-approved medications (e.g., opioids, benzodiazepines, antipsychotics, antiepileptics, antihistamines, etc.) that may interfere with memory (Budson & Solomon, 2011), particularly those that have strong anticholinergic effects, such as tricyclic antidepressants and certain mood stabilizers (Julien, 2007). Given the likelihood that about 90% of older Americans have taken at least one prescription medication in the past month (Gu, Dillon, & Burt, 2010), the likelihood that one or more of those medications may have some effect on cognition is high (Obermann, Morris, & Roe, 2013). However, because the results of this study were intended to be generalized to a population whose medication history is unknown at the time of testing, participants taking one or more FDA-approved medications were excluded from the analysis only if the evaluation resulted in a recommendation that a medication be discontinued due to its adverse cognitive effects.

A variety of neurodegenerative disorders fall under the umbrella term *dementia*, but there is heterogeneity in the presentation and symptoms depending on presumed etiology. Therefore, participants diagnosed with other neurocognitive disorders (including but not limited to vascular dementia, mixed dementia, dementia with Lewy bodies, frontotemporal dementia, Parkinson's

disease, corticobasal degeneration, progressive supranuclear palsy, normal pressure hydrocephalus, Wernicke-Korsakoff syndrome, or variant Creutzfeldt-Jakob disease) were excluded. In order to minimize the influence that the residual cognitive symptoms of a brain injury may have on test performance, participants were excluded if they (or knowledgeable informants such as a caregivers or close relatives) reported a history of traumatic brain injury with loss of consciousness greater than 5 seconds within 5 years of the time memory symptoms were first noticed. Participants with a documented metabolic disorder, such as hypothyroidism or vitamin B_{12} deficiency, were excluded, as such deficiencies can mimic dementias and/or exacerbate cognitive symptoms. Finally, because people with Down syndrome almost invariably develop AD pathology (Budson & Solomon, 2011), one patient with Down syndrome was excluded from the sample.

Depressive symptoms can cause or exacerbate cognitive symptoms, and the distinction between the two is not always clear. The onset of later-life depression often co-occurs with the onset of cognitive symptoms, and apathy is often a behavioral symptom of AD (Budson & Solomon, 2011). Therefore, patients who are being treated for depression will be included unless it is clear, as judged by the diagnostic team in reviewing the results of a depression screening measure (Geriatric Depression Scale) and diagnostic interview, that major depressive disorder is the primary cause of the participant's reported cognitive symptoms, in the absence of a diagnosis of AD or MCI.

Demographic characteristics. The overall sample consisted of 40 participants, of which 18 (45%) were male and 22 (55%) were female. Age ranged from 58-90, level of education ranged from 8-20, and estimated premorbid intellectual ability (WTAR Estimated IQ) ranged

from 82-119. Table 2 lists means and standard deviations for demographic variables in each group.

Table 2

Demographic Characteristics by Group

	Clinical Group			Control Group			Mean		
Variable	Mean	SD	*n*	Mean	SD	*n*	Difference	*p*	*d*
Age	79.4	7.5	20	73.3	6.9	20	6.1	.011	.85
Years of Education	14.6	3.2	20	16.9	2.8	20	-2.3	.018	-.76
Estimated Premorbid IQ (WTAR)	104.6	12.8	20	113.3	5.9	20	-8.7	.010	-.87

Note. WTAR = Wechsler Test of Adult Reading.

Clinical group. Out of 79 patients who presented for clinical evaluation, 28 were excluded (5 previous head injury, 5 ongoing medical condition, 5 psychiatric, 4 prior stroke, 4 presenting problem not related to memory, 2 previously diagnosed, 1 Down syndrome, 1 blind, 1 deaf). Of the remaining 51 patients who were offered an opportunity to participate, 11 declined, leaving 24 participants in the preliminary clinical group. After all participants in the clinical group completed the research battery, 4 participants were excluded from analyses due to receiving a diagnosis that was not AD or MCI (depression/anxiety = 3, mixed dementia = 1). The final clinical group was comprised of 20 participants, of which 14 were diagnosed with AD and

6 with MCI. Of those 20 participants, 8 were male (40%) and 12 were female (60%). Age ranged from 69-90, level of education ranged from 8-20, and estimated premorbid IQ ranged from 82-119. Table 2 lists means and standard deviations for demographic variables in the clinical group. In the clinical group, the mean MMSE score was 21.9 (*SD* = 4.6), with a range of 13-27. The mean MoCA score was 17.2 (*SD* = 4.8), with a range of 9-24. The mean ADAS-Cog score was 15.9 (*SD* = 8.5), with a range of 5-35. The mean ADL score was 24.1% (*SD* = 22.9%), with a range of 1%-69%. All patients lived in private residences, either independently or with an unpaid caregiver (typically a spouse, son, or daughter).

Control group. All potential control group participants consented to participate in the study. The control group consisted of 20 participants, of whom 10 were male (50%) and 10 were female (50%). Table 2 lists means and standard deviations for demographic variables in the control group. Of note, the level of education of the control group differed dramatically from population estimates derived from U.S. Census data from 2010. Of Americans over the age of 60, 34.78% held at least a high school diploma, 13.47% held at least a bachelor's degree, 5.40% held at least a master's degree, and 1.02% held at least a doctorate or professional degree. In the control group, all participants (100%) held at least a high school diploma, 15 participants (75%) held at least a bachelor's degree, 11 participants (55%) held at least a master's degree, and 5 participants (25%) held at least a doctorate. A review conducted by Sinnett and Holen (1999) found that WMS-R VR I is sensitive to age and VRII is sensitive to age, SES, and age × SES interaction.

Comparisons between groups. Independent-samples *t*-tests were carried out to determine whether there were meaningful differences in demographic characteristics between the

clinical and control groups. The mean age of the clinical group was significantly greater than the mean age of the control group ($t(38) = -2.66$, $p = .011$, $d = -.85$). The mean level of education of the clinical group was significantly smaller than the mean level of education of the control group ($t(38) = 2.47$, $p = .018$, $d = .77$). The mean estimated premorbid IQ of the clinical group was significantly less than the mean estimated premorbid IQ of the control group, $t(38) = 2.763$, $p = .010$, $d = -.87$. A chi-square test did not reveal a significant difference in gender between groups ($\chi^2 = .404$, $p = .751$). Differences between groups are summarized in Table 2 above.

In the overall sample, age, education, and estimated premorbid IQ were all significantly correlated with each other. Age was moderately negatively correlated to level of education ($r = -.418$, $p = .007$), and to estimated premorbid IQ ($r = -.331$, $p = .037$). Level of education was strongly positively correlated to estimated premorbid IQ ($r = .560$, $p < .001$). Within each group, some correlations between demographic variables were significant (Table 3).

Table 3

Correlations Between Demographic Variables by Group

Subtest	Age	Years of Education	Estimated Premorbid IQ
Age		-.369	**-.498**
Education	-.276		**.557**
Estimated Premorbid IQ	-.090	**.479**	

Note. Unshaded = clinical group, shaded = control group; bold text indicates significance at or below the $p = .05$ level; WTAR = Wechsler Test of Adult Reading.

Measures

The three measures were chosen for their ease of use, their uniqueness, and the short amount of time required to administer. For the WMS-IV and the BVMT-R, participants were told to expect a delayed recall condition. They did not expect a delayed recognition condition for WRAML2 subtests.

Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2). The Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2; Sheslow & Adams, 2003) is an individually-administered, standardized battery of memory tests for use in children and adults ages 6 to 89. The Design Memory and Picture Memory subtests, along with their recognition components, were used in this study. The Design Memory subtest consists of 5 cards with increasingly complex abstract designs. For each card, the participant is allowed 5 seconds to study the designs before the card is removed from view. After a 10-second delay, the participant is asked to draw the designs from memory with as much detail as possible on a response form. There are 12 possible correct responses for each card; raw scores for the entire subtest range from 0-60. The Picture Memory subtest consists of 4 full-color drawings of common scenes (e.g., a family at a zoo, some friends watching television). For each card, the participant is allowed 10 seconds to study the drawing before the card is removed from view. Another drawing is then presented that is similar to the original with certain key components moved, changed, or added. For each repetition, the participant must place an 'X' on any of the 8 parts of each scene that has been moved, changed, or added. Raw scores for the entire subtest range from 0-32. Design Memory Recognition is a yes-no recognition task consisting of 46 designs, 23 of which were present on the original cards. Picture Memory recognition is a yes-no recognition task

consisting of 44 smaller sections of the original 4 drawings or the 4 modified drawings, some of which were present and some of which are new. Both are administered approximately 20 minutes after the completion of the respective immediate recall subtests. For each recognition subtest, the participant must indicate whether a particular part of the design or drawing was seen during the immediate recall subtest; there are 46 items on the Design Memory Recognition subtest and 44 items on the Picture Memory Recognition subtest. Raw scores for the Design Memory Recognition subtest range from 0-46; however, scores of less than 23 are highly unlikely, as a response of "No" on every item (the participant does not recall any of the true or sham items) would yield a score of 23. Raw scores for the Picture Memory Recognition subtest range from 0-44; scores of less than 22 are highly unlikely, as a response of "No" on every item (the participant does not recall any of the true or sham items) would yield a score of 22. For both recognition subtests, random guessing should yield a score of at least 50% correct.

The WRAML2 demonstrated adequate reliability and validity in normative studies. Coefficient alphas for the subtests and age groups used in this study are shown in Table X below. Inter-rater reliability using the rubrics provided to score the Design Memory subtest ranged from .976 to .981 (Sheslow & Adams, 2003). In clinical samples, both Design Memory and Picture Memory demonstrated utility in distinguishing AD patients from matched controls (Table 4).

Wechsler Memory Scale, 4th Edition (WMS-IV). The Wechsler Memory Scale, 4th Edition (WMS-IV; Wechsler, 2009) is an individually-administered, standardized battery of memory tests. Only the Visual Reproduction subtests, including Visual Reproduction Recognition, were used in the study. This subtest consists of 5 cards on which increasingly complex geometric designs are printed. There are 2 designs on each of the final two cards. Each

Table 4

Scaled Score Differences between AD Clinical Sample and Matched Controls on Selected WRAML2 Subtests

	AD			Controls			Mean		
Subtest/Index	Mean	SD	*N*	Mean	SD	*N*	Difference	*p*	*r*
Design Memory	6.9	3.6	17	10.1	3.5	17	-3.1	.015	0.89
Picture Memory	6.1	3.4	16	8.9	3.1	17	-2.8	.020	0.85
Design Memory Recognition	9.1	4.0	17	11.2	2.9	17	-2.1	.086	.61
Picture Memory Recognition	7.9	3.7	17	9.1	2.5	17	-1.1	.315	.35

Note. AD = Alzheimer's disease; WRAML2 = Wide Range Assessment of Memory and Learning, 2nd Edition. Source: Sheslow & Adams (2003).

card is presented individually. The participant is allowed 10 seconds to study the card before it is removed from view. The participant must then draw the design(s) as accurately as possible on a blank response form. Responses are scored according to criteria provided in the test manual. After a delay of 20-30 minutes, the participant is given a blank response form and asked to reproduce the design (or designs) seen on any 1 of the 5 cards that were presented earlier. Once the participant is satisfied with the reproduction, another blank response form is presented and the participant is asked to reproduce another of the 5 cards. Three more blank response forms are

presented one at a time until all cards have been attempted or until the participant is unable to recall any more of the designs. Raw scores on each task (immediate and delayed recall) range from 0-43. Following the delayed recall task, there is a multiple-choice recognition task. The participant is presented with 7 cards, one at a time. Each card has 6 designs on it, 5 of which look similar to a design that was presented earlier and one of which is exactly the same. The participant is asked to identify which of the 6 designs is the one that was presented earlier. Raw scores on this task range from 0-7.

The WMS-IV demonstrated adequate reliability and validity in normative studies. Coefficient alphas ranged from .92 to .94 for Visual Reproduction I (VR I) and from .95 to .97 for Visual Reproduction II (VR II) in the 65- to 90-year-old age group. WMS-IV Visual Reproduction demonstrated utility in distinguishing between AD/MCI and normal controls (Table X). In the AD clinical sample, reliability coefficients were .96 for VR I and .99 for VR II. In the MCI clinical sample, reliability coefficients were .93 for VR I and .97 for VR II. WMS-IV Visual Reproduction demonstrated utility in distinguishing between AD/MCI and normal controls (Table 5).

Brief Visuospatial Memory Test, Revised Edition (BVMT-R). The Brief Visuospatial Memory Test, Revised (BVMT-R; Benedict, 1997) is an individually administered, standardized test of visual memory and learning. It was developed as a non-verbal companion to the Hopkins Visual Learning Test (HVLT; Benedict & Groninger, 1995). The examinee is shown a letter-sized (8.5 × 11 inch) card on which six simple designs are printed in a 2 × 3 arrangement. The examinee is allowed look at the card for 10 seconds, after which the card is removed from view and the participant is asked to draw as many of the designs as he or she can remember, in the

Table 5

Scaled Score Differences between AD Clinical Sample and Matched Controls on Selected WMS-IV Subtests

	Clinical Group		Normal Controls				
Subtest	Mean	SD	Mean	SD	*N*	Mean Difference	*p*
MCI							
VR I	8.7	3.0	10.7	2.6	50	1.92	< .01
VR II	8.0	3.9	10.1	2.7	49	2.12	< .01
AD							
VR I	5.6	3.4	10.8	2.6	48	5.19	< .01
VR II	4.2	3.1	10.1	3.1	48	5.96	< .01

Note. MCI = mild cognitive impairment, AD = Alzheimer's disease, WMS-IV = Wechsler Memory Scale, 4th Edition, VR I = Visual Reproduction I, VR II = Visual Reproduction II. Source: Wechsler (2009).

same location on the page. The process is repeated with the same stimulus card for two more trials. The BVMT-R is not as widely used as some of the other measures available (e.g., the WMS-IV and the WRAML2), but it generates some additional scores that are not available when using other measures. There are three learning trials, a 25-minute delayed recall trial, and a delayed recognition trial that is administered immediately after the delayed recall trial. In addition to those five trials, scores are generated for Learning (Trial 2 or Trial 3, whichever is

greater, minus Trial 1), Total (sum of Trials 1-3), Percent Retained (delayed recall divided by the greater score of Trial 2 or Trial 3), Recognition Hits (number of true positives) and Recognition False Alarms (number of false positives), Discrimination Index (Recognition Hits minus Recognition False Alarms), and Response Bias, a derived score that represents a bias toward positive or negative responses during the recognition trial (scores range from 0 to 1, with higher scores suggesting a bias toward "yes" responses that leads to a high number of false positives). An advantage of the BVMT-R, particularly when assessing decline over 3- or 6-month intervals, is the availability of six alternate forms. The authors of the BVMT-R claim that Form 2 shows the strongest test-retest reliability characteristics, and this is the form that was used in the study.

In normative studies, BVMT-R scores were strongly correlated with other visual memory measures, including the Wechsler Memory Scale, Revised Edition (WMS-R) and the Rey-Osterrieth Complex Figure Test (RCFT), suggesting adequate convergent and divergent validity (Table 6). In clinical samples, the BVMT-R demonstrated utility in distinguishing between AD/mixed dementia groups and normal controls, suggesting adequate discriminative validity (Table 7).

When working with AD and MCI patients, clinicians must measure multiple domains in a short amount of time, ideally in one test session, with patients who are easily fatigued and are likely to be cognitively impaired. The battery used at The Memory Clinic consists of selected subtests and has remained relatively unchanged (aside from the addition of new versions of the same subtests) for 25 years. One of the proposed measure (Visual Reproduction from the WMS-IV) is part of that battery. The other three (WRAML2 Design Memory and Picture Memory, and

Table 6

Correlations between BVMT-R Scores and Scores on Similar Visual Memory Measures

Measure	CFT-C	CFT-R	VR I	VR II
Immediate Recall Total	.66	.78	.66	--
Delayed Recall	.65	.78	--	.80
Discrimination Index	.33	.52	--	.50
Recognition Response Bias	-.08	.00	--	.09

Note. BVMT-R = Brief Visuospatial Memory Test, Revised Edition; CFT-C = Copy trial of the Rey-Osterrieth Complex Figure Test; CFT-R = Recall trial of the Rey-Osterrieth Complex Figure Test; VR I = Visual Reproduction I subtest of the Wechsler Memory Scale, Revised Edition (WMS-R); VR II = Visual Reproduction II subtest of the WMS-R.
Source: Benedict, Schretlen, Groninger, & Dobraski (1996).

the BVMT-R) were chosen for their ease of use, their uniqueness in terms of item content and administration procedures, and the short amount of time required to administer and score them.

The BVMT-R is unique among measures used in this study because of the use of multiple trials to assess learning with repetition. This is important; for example, Greene, Baddeley, and Hodges (1996) administered the doors and people test, a test of visual and verbal recall and recognition, to a sample of 33 AD patients and 30 matched normal controls. They found that the AD group performed worse than the control group on the doors and people test as well as on a verbal memory measure. Performance of the AD group on the doors and people test indicated that impaired learning (i.e., encoding difficulty) was a hallmark of the AD group. While there was a difference between groups in the rate of forgetting verbal material, this difference was not

Table 7

Mean Differences Between AD and Mixed Dementia Clinical Samples and Normal Controls for Selected BVMT-R Scores

BVMT-R Score	Normal Controls	Alzheimer's Dementia	Vascular or Mixed Dementia	F or χ^2	p
Immediate Recall (Total)					
M	19.8	6.2	6.4	132.3	< .0001
SD	6.2	3.5	2.9		
Delayed Recall					
M	7.6	1.5	2.0	133.9	< .0001
SD	2.7	1.5	1.3		
Recognition Hits					
M	5.7	4.8	4.5	23.3	< .0001
SD	0.6	1.3	1.5		
Recognition False Alarms					
M	0.1	1.6	1.1	44.7	< .0001
SD	0.4	1.5	1.3		

Note. BVMT-R = Brief Visuospatial Memory Test, Revised Edition. Source: Benedict (1997).

observed for visual material. Results suggest that a multiple-trial visual memory measure make important contributions to the assessment of episodic memory impairment in AD that are not provided by single-trial measures. Table 8 describes the various measures used in this study in terms of time, item content, and methodology.

Table 8

Comparison of Time Points, Item Content, and Retrieval Paradigm between Visual Memory Measures Used in the Present Study

Subtest	Time Point	Item Content	Retrieval Paradigm
WRAML2 Design Memory	Immediate	Geometric designs	Recall
WRAML2 Picture Memory	Immediate	Visual scenes	Recognition
WMS-IV Visual Reproduction I	Immediate	Geometric designs	Recall
BVMT-R Immediate Recall	Immediate	Geometric designs in specific locations	Recall
WRAML2 Design Memory Recognition	Delayed	Parts of geometric designs	Yes-No Recognition
WRAML2 Picture Memory Recognition	Delayed	Parts of visual scenes	Yes-No Recognition
WMS-IV Visual Reproduction II	Delayed	Geometric designs	Recall
BVMT-R Delayed Recall	Delayed	Geometric designs in specific locations	Recall
BVMT-R Recognition	Delayed	Individual geometric designs	Yes-No Recognition

Note. WRAML2 = Wide Range Assessment of Memory and Learning, 2nd Edition; WMS-IV = Wechsler Memory Scale, 4th Edition; BVMT-R = Brief Visuospatial Memory Test, Revised Edition.

Procedure

Study procedures were reviewed and approved by the George Fox University Human Subjects Research Committee before data collection commenced. As part of The Memory

Clinic's routine policies, each participant and/or a legally-authorized representative (LAR) signed a form consenting to the requirements of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Each participant and/or the LAR then signed a form consenting to participate in the study (Appendix D).

Study measures were administered as part of the new-patient intake evaluation. Briefly, the diagnostic process used at The Memory Clinic proceeds as follows. On the first visit, the patient meets with a psychometrist (either a doctoral intern or a post-doctoral fellow) to complete the neuropsychological test battery. This takes approximately 1.5 to 2 hours. In the meantime, a doctoral-level psychologist reviews the patient's history with caregivers and/or other reliable informants using a semi-structured interview format. When testing is completed, a phlebotomist draws blood from the patient for off-site laboratory analysis. The patient and his/her family then reconvene with the psychometrist and the psychologist to complete the interview with the patient and to discuss next steps. A follow-up visit is scheduled for 3-4 weeks later. Between the first and second visit, the patient may undergo magnetic resonance imaging (or X-ray computed tomography if indicated, e.g., for patients with pacemakers). If the case is diagnostically complex, an appointment with a neurologist may be scheduled. At the follow-up visit (or after the neurology visit), the patient is seen by a geriatrician who reviews medical history, laboratory and imaging findings and completes a brief physical and neurological examination (if this was not done by the neurologist). A summary of the medical evaluation is provided to the psychologist, who then meets with the patient and caregivers and reviews the results of the evaluation, explains the diagnostic picture, and discusses treatment options. Patients are typically

scheduled for follow up in 3-6 months in order to document and track any changes in cognition and function over time.

With the addition of the visual memory measures, each new patient test battery took between 2 and 2.5 hours to complete. Each test battery was completed in a single visit without interruption. All study measures were administered in a clean, well-lit facility, in accordance with standard administration procedures delineated in the administration manuals. All measures were administered by examiners who were already trained in standardized test administration and adhered to the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct (2002). The examiners were doctoral interns with Master's degrees in clinical psychology who were enrolled in APA-accredited doctoral programs in clinical psychology. Examiners were supervised by two doctoral-level clinical neuropsychologists. Participants were provided with opportunities to take breaks at any time in order to minimize fatigue; when breaks were anticipated to interfere with delayed recall time limits, they were postponed until after the delayed recall measure was completed.

In order to minimize order-of-administration effects without rearranging the entire new patient test battery, only the visual memory measures were counterbalanced within the battery. In this way, each participant completed non-visual memory measures in the same order, with one or another visual memory measure inserted at one of 3 points in the sequence. Each participant may completed the three visual memory measures within the new patient test battery in one of 6 possible orders.

Patients who present for memory evaluation sometimes become confused as the number of tests increases and as fatigue increases. When asked to recall information after a delay during

which several other distracting activities may have taken place, some patients may lose track of what they are being asked to recall. In order to minimize this type of confusion, modifications were made to the stimulus cards and response forms for each measure (with the exception of WRAML2 Picture Memory). Stimulus cards and response forms for each measure were photocopied and re-printed on paper of different colors, allowing a specific measure color to be identified later by color (e.g., “Please draw what you saw on the green cards that I showed you a few minutes ago.”) This minor deviation from standard procedures was not expected to have any influence on the validity of results. Observational evidence indicated that, for those participants who became confused about what material they were being asked to recall, the color-coding strategy did not provide much assistance, and may not have been necessary.

Chapter 3

Results

All statistical procedures, with the exception of ROC curve analyses, were completed using Statistical Package for the Social Sciences (SPSS) software. Subtest raw scores were converted to scaled scores ($M = 10$, $SD = 3$) using normative data provided in the respective tests' administration manuals, with the exception of the BVMT-R. Normative data for the BVMT-R are provided as T scores ($M = 50$, $SD = 10$), which were converted to z scores ($M = 0$, $SD = 1$) using the formula $z = (T - 50) / 10$ and then to scaled scores using the formula $SS = 3z + 10$. However, BVMT-R normative data are only provided up to and including age 79. Therefore, BVMT-R raw scores for participants aged 80 or older were converted to z scores using means and standard deviations provided by Gale and colleagues (2007) and from z scores to standard scores using the formula $SS = 3z + 10$.

Comparison of Groups' Performance to Each Other and to Normative Data

In order to identify similarities and differences between the participants in the present study and participants in normative studies, thereby allowing conclusions to be made about generalizability of the data, mean scaled scores and correlations between measures were compared to normative data provided in the respective test manuals. One-sample t-tests were performed using a population mean of 10, using a Bonferroni correction for multiple comparisons to minimize Type I error. In the control group, mean scaled scores on several measures (specifically, the WRAML2 recognition tasks and the immediate recall tasks of the

WMS-IV and the BVMT-R) were greater than the population mean scaled score of 10. Clinical and control group means, *SD*s and significance levels comparisons between the groups are found in Table 9. The mean scaled score was lower for the clinical group than for the control group on every measure. Comparisons between the control group and the population means (and SDs) are found in Table 10. In the clinical group, all means were significantly lower than the population mean scaled score of 10, with the exception of WRAML2 Picture Memory Recognition, which approached but did not reach significance ($p = 0.055$). Comparisons between the clinical group and the population means (and SDs) are found in Table 11. In the clinical group, mean scaled scores for every measure except WRAML2 Picture Memory Recognition were significantly lower than the population mean scaled score of 10.

For comparison purposes, the inter-correlations between measures for each group and the respective standardization samples are found in Tables 12 and 13. As can be seen in Table 12, correlations for the control group ranged from -.18 to .68, compared to the standardization samples, whose correlations ranged from .26 to .78. The average correlation of the control group was .259 (*SD* = .359), and for the standardization group it was .474 (*SD* = .197). Correlations were compared using a Fisher transformation to convert correlation coefficients into standard *z* scores. Computed *z* scores were then compared to each other using the formula [z_1 - z_2 / sqrt (s_1 + s_2)], which divides the difference between means by the pooled variance of the two groups. Table 12 demonstrates that differences emerged between some measures but not others. Specifically, the correlation between Picture Memory and Picture Memory Recognition, and between Visual Reproduction I and Visual reproduction II were lower in the control group than those reported in the measures' respective manuals. Similarly, as can be seen in Table 13,

Table 9

Mean Scaled Score Differences Between Clinical Group and Control Group

	Clinical Group		Control Group				
Subtest/Index	Mean	*SD*	Mean	*SD*	Mean Difference	*p*	*d*
Design Memory	5.7	2.2	10.2	2.7	-4.5	**< .001**	-1.83
Picture Memory	8.0	2.6	11.1	2.3	-3.1	**< .001**	-1.26
Design Memory Recognition	8.2	2.6	11.5	2.6	-3.3	**< .001**	-1.27
Picture Memory Recognition	8.7	2.8	11.5	1.5	-2.8	**< .001**	--
WMS-IV Visual Reproduction I	5.6	3.6	12.1	3.8	-6.5	**< .001**	-1.76
WMS-IV Visual Reproduction II	3.8	2.1	10.2	2.5	-6.4	**< .001**	-2.77
BVMT-R Immediate	4.9	3.6	12.0	3.5	-7.1	**< .001**	-2.00
BVMT-R Delayed	4.0	2.5	11.2	4.4	-7.2	**< .001**	--

Note. Bold text represents correlation coefficients that are significant at or below the $p = .05$ level. WRAML2 = Wide Range Assessment of Memory and Learning, 2nd Edition; WMS-IV = Wechsler Memory Scale, 4th Edition; BVMT-R = Brief Visuospatial Memory Test, Revised Edition.

correlations for the clinical group ranged from -.17 to .52. The average correlation of the clinical group was .137 (*SD* = .209). Four of the 9 correlations differed significantly between the control group and standardization samples, while only 3 differed significantly between the clinical group and standardization samples. The correlations between VR I and VR II and between VR II and

Table 10

Mean Differences between Control Group and Population Mean of 10 on Selected Visual Memory Measures

Measure	Mean	*SD*	*N*	Mean Difference	*t*	*p*	*d*
DM	10.20	2.71	20	0.20	0.330	.745	.07
PM	11.05	2.31	20	1.05	2.037	.056	.37
DM Recognition	11.45	2.56	20	1.45	2.529	**.020**	.52
PM Recognition	11.45	1.50	20	1.45	4.313	***<.001***	.61
VR I	12.05	3.85	20	2.05	2.384	**.028**	.59
VR II	10.15	2.46	20	0.15	0.273	.788	.05
BVMT-R Immediate	12.00	3.48	20	2.00	2.568	**.019**	.62
BVMT-R Delayed	11.23	4.37	20	1.23	1.254	.225	.33

Note. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, 4th Edition (WMS-IV) Visual Reproduction I; VR II = WMS-IV Visual Reproduction II. Bold text represents *p* levels that are significant at or below the .05 level; bold and italic text represents *p* levels that are significant at or below the Bonferroni-corrected level of $p = .006$.

BVMT-R Delayed Recall were significantly lower in both groups than in standardization samples.

Comparison of Clinical Group Performance to similar Clinical Samples

Independent-samples *t*-tests were conducted to compare this study's clinical group demographics to those of similar clinical groups reported in each measure's respective test manual. Analyses did not reveal a significant difference between the mean age of the clinical group and that of the WRAML2 clinical group ($t(35) = -0.8824$, $p = .384$, $d = -.291$). A chi-

Table 11

Mean Differences between Clinical Group and Population Mean of 10 on Selected Visual Memory Measures

Measure	Mean	*SD*	*N*	Mean Difference	*t*	*p*	*d*
DM	5.70	2.23	20	-4.30	-8.636	**< .001**	-1.63
PM	8.00	2.62	20	-2.00	-3.419	**0.003**	-.71
DM Recognition	8.20	2.59	20	-1.80	-3.111	**0.006**	-.64
PM Recognition	8.70	2.85	20	-1.30	-2.041	0.055	-.44
VR I	5.60	3.65	20	-4.40	-5.395	**< .001**	-1.32
VR II	3.80	2.07	20	-6.20	-13.412	**< .001**	-2.41
BVMT-R Immediate	4.89	3.63	20	-5.11	-6.294	**< .001**	-1.53
BVMT-R Delayed	3.99	2.51	20	-6.02	-10.726	**< .001**	-2.17

Note. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, 4th Edition (WMS-IV) Visual Reproduction I; VR II = WMS-IV Visual Reproduction II. Bold text represents *p* that are significant at or below the .05 level; bold and italic text represents correlations that are significant at or below the Bonferroni-corrected level of $p = .006$.

square test did not reveal a significant difference in gender between the two samples ($\chi^2 = 2.2452$, $p = .134$). The clinical group had a significantly greater level of education than the WRAML2 clinical group (Fisher's Exact Test, $p = .0164$). Analyses did not reveal a significant difference between the mean age of the clinical group and the BVMT-R clinical group ($t(59) = 1.9064$, $p = .062$, $d = .50$). The mean level of education of the clinical group was significantly greater than that of the BVMT-R clinical group ($t(59) = 17.2600$, $p < .001$, $d = 1.11$). There was

Table 12

Comparison of Correlation Coefficients between Control Group and Normative Data Reported by Test Authors

Measure 1	Measure 2	$r_{standard}$	$r_{control}$	z	p
DM	PM	.41	.58	-0.929	.176
	DM Recognition	.54	.66	-0.772	.220
	PM Recognition	.29	-.18	1.967	**.025**
PM	DM Recognition	.26	.51	-1.214	.112
	PM Recognition	.39	-.16	2.347	**.010**
DM Recognition	PM Recognition	.26	-.12	1.600	0.055
VR I	VR II	.66	.19	2.435	**.008**
	BVMT-R Immediate	.68	.68	0.000	1.00
VR II	BVMT-R Delayed	.78	.17	3.542	**< .001**

Note. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, Revised Edition (WMS-R) Visual Reproduction I; VR II = WMS-R Visual Reproduction II. Bold text represents correlation coefficients that are significant at or below the $p = .05$ level. WRAML2 inter-correlations are derived from a sample ranging in age from 9 to 89 years.

no significant difference between mean age of the clinical group and the WMS-IV AD clinical group ($t(66) = 0.4752$, $p = .636$, $d = .12$). A chi-square test did not reveal a significant difference in gender between the two samples ($\chi^2 = 0.1276$, $p = .721$). There was no significant difference in level of education between the two samples (Fisher's Exact Test, $p = .942$). The mean age of the clinical group was significantly greater than that of the WMS-IV MCI clinical group ($t(68) = 3.404$, $p = .001$, $d = .89$). A chi-square test did not reveal a significant difference in gender between the two samples ($\chi^2 = 1.12$, $p = .290$). There was no significant difference in level of

Table 13

Comparison of Correlation Coefficients between Clinical Group and Normative Data Reported by Test Authors

Measure 1	Measure 2	$r_{standard}$	$r_{clinical}$	z	p
DM	PM	.41	-.17	2.480	**.007**
	DM Recognition	.54	.22	1.554	.060
	PM Recognition	.29	-.02	1.301	.097
PM	DM Recognition	.26	.08	0.759	.224
	PM Recognition	.39	.34	0.236	.407
DM Recognition	PM Recognition	.26	-.03	1.209	.113
VR I	VR II	.66	.09	2.848	**.002**
	BVMT-R Immediate	.68	.52	1.025	.153
VR II	BVMT-R Delayed	.78	.20	3.416	**< .001**

Note. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, Revised Edition (WMS-R) Visual Reproduction I; VR II = WMS-R Visual Reproduction II. Bold text represents correlation coefficients that are significant at or below the $p = .05$ level. WRAML2 inter-correlations are derived from a sample ranging in age from 9 to 89 years.

education between the two samples (Fisher's Exact Test, $p = .842$). In the clinical group, all means differed significantly from the standardization mean scaled score of 10. Tables 14, 15, and 16 compare means and *SD*s between this study's clinical group and the clinical groups described in each measure's respective test manual. In summary, the clinical group in this study was more highly educated than those conducted as part of the standardization studies for the WRAML2, the BVMT-R, and the WMS-IV MCI group, while there was no difference in education when compared to the AD group in the WMS-IV clinical study. No differences in age or gender were

found between this study's clinical group and the standardization clinical groups, with the exception of the WMS-IV MCI group.

Table 14

Clinical Group Compared to AD Clinical Group reported in the WRAML2 Test Manual

	WRAML2								
	Clinical Group			Clinical Group					
Subtest/Index	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean Difference	*p*	*d*
Design Memory	5.7	2.2	20	6.9	3.6	17	-1.2	.222	-.40
Picture Memory	8.0	2.6	20	6.1	3.4	16	1.9	.066	.63

Note. AD = Alzheimer's disease; WRAML2 = Wide Range Assessment of Memory and Learning, 2nd Edition.

Normative data from the WMS-IV *Technical and Interpretive Manual* (Wechsler, 2009) were provided using separate groups for AD and MCI and are found in Table 15. The similarity of demographic data and mean scaled scores between the clinical group in the current study and the two WMS-IV clinical groups suggest that this current study's clinical group more closely resembles the WMS-IV AD clinical group than the MCI clinical group.

Overall, these comparative results indicate that the performance of the control group differed fairly consistently from standardization group data reported by the tests' authors. Specifically, mean scaled scores for most measures were higher in this study's control group than in most standardization samples. Results generated using control group data should

Table 15

Clinical Group Compared to MCI and AD Clinical Groups reported in the WMS-IV Test Manual

	Clinical Group (MCI *and* AD)			WMS-IV Clinical Group (MCI *or* AD)			Mean		
Subtest/Index	Mean	*SD*	*n*	Mean	*SD*	*n*	Difference	*p*	*d*
MCI Group*									
VR I	5.6	3.6	20	8.7	3.0	50	-3.1	<.001	-.94
VR II	3.8	2.1	20	8.0	3.9	49	-4.2	<.001	-1.34
AD Group*									
VR I	5.6	3.6	20	5.6	3.4	48	0.0	1.00	0.00
VR II	3.8	2.1	20	4.2	3.1	48	-0.4	.600	-.15

Note. MCI = mild cognitive impairment, AD = Alzheimer's disease, WMS-IV = Wechsler Memory Scale, 4th Edition, VR I = Visual Reproduction I, VR II = Visual Reproduction II.

* Because this study used a combined MCI/AD clinical group, comparisons were made between this study's clinical group and each of the two clinical groups (MCI and AD) reported separately in the WMS-IV test manual.

therefore be interpreted with caution, as it may not represent a typical population of non-impaired individuals. However, for the clinical group used in the current study, performance on measures of visual memory was found reasonably similar to that of comparable clinical samples reported in the respective test manuals. Of the eight comparisons made, two comparisons indicated that the clinical group in the present study was less impaired than the clinical group studied by the measure's authors. Two comparisons found an equivalent performance between clinical groups. Four comparisons demonstrated that the clinical group in the present study was

Table 16

Clinical Group Compared to AD Clinical Group reported in the BVMT-R Test Manual

	Clinical Group			BVMT-R Clinical Group					
Subtest/Index	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean Difference	*p*	*d*
Total Recall	4.9	3.6	20	6.2	3.5	41	-1.3	.182	-.37
Delayed Recall	4.0	2.5	20	1.5	1.5	41	2.5	<.001	1.21

Note. BVMT-R = Brief Visuospatial Memory Test, Revised Edition.

more impaired. However, two of these comparisons were between an MCI group (WMS-IV) and the present study's clinical group. Ignoring these two comparisons, no differences were significant with the exception of BVMT-R Delayed Recall. Performance on this measure was significantly greater in the present study's clinical group than in the clinical group studied by the author. Results suggest that the performance of the clinical group on selected memory measures in this study is similar to that of individuals with memory impairment due to AD in other clinical samples. The clinical group in this study is representative of the typical population of individuals who undergo memory testing and are found to have AD. Comparisons between the performance of this study's clinical group and groups described in the respective tests' manuals can therefore be interpreted and generalized to this population.

Hypotheses 1 and 2: Relatedness Among Visual Memory Measures

As found in Table 17, correlations between measures were higher overall in the control group (M = .289, SD = .350) than in the clinical group (M = .157, SD = .299). Eleven of the 28

Table 17

Intercorrelations of Scaled Score Performance for the Control and Clinical Groups

Subtest	1	2	3	4	5	6	7	8
1. WRAML2 Design Memory		***.580***	***.661***	-.178	.328	**.526**	**.508**	.122
2. WRAML2 Picture Memory	-.172		.512	-.159	**.534**	**.538**	***.647***	.435
3. WRAML2 Design Memory Recognition	.221	.078		-.124	.227	.189	***.647***	.262
4. WRAML2 Picture Memory Recognition	-.023	.339	-.034		-.177	-.190	-.269	-.421
5. WMS-IV Visual Reproduction I	***.749***	-.221	.316	.008		.189	***.677***	**.711**
6. WMS-IV Visual Reproduction II	-.073	.282	-.061	.141	.094		.329	.171
7. BVMT-R Total Recall	***.426***	-.115	.074	-.335	**.516**	.117		***.825***
8. BVMT-R Delayed Recall	***.512***	-.026	.273	-.271	**.518**	.204	***.871***	

Note. Unshaded = clinical group, shaded = control group; bold text indicates significance at or less than the $p = .05$ level; bold and italic text indicates significance at or less than the $p = 0.002$ level. WRAML2 = Wide Range Assessment of Memory and Learning, 2nd Edition; WMS-IV = Wechsler Memory Scale, 4th Edition; BVMT-R = Brief Visuospatial Memory Test, Revised Edition.

correlations were significant for the control group, while only six were significant for the clinical group. Given the moderate to strong correlations between demographic variables and scaled scores shown in Table 18, these values may be misleading. Relationships between measures appear to be strongly influenced by demographic variables, and it is not clear from this table the extent to which the measures are related on their own. To control for the influence of age, education, and IQ, partial correlations were calculated for each group. The obtained partial correlations are shown in Table 19. In the control group, partial correlations between measures

became stronger after controlling for demographic variables. This is likely due to the homogeneity of the sample in terms of level of education and estimated premorbid IQ.

Table 18

Correlations between Visual Memory Mean Scaled Scores and Demographic Variables by Group

Group	Overall			Clinical			Control		
Measure	Age	Ed	IQ	Age	Ed	IQ	Age	Ed	IQ
DM	**-.385**	**.366**	**.518**	.013	.186	.307	-.338	.150	**.559**
PM	**-.402**	.244	.261	.065	.019	-.024	**-.624**	.108	.255
DM Recognition	**-.372**	.247	.214	-.288	.182	-.037	-.112	-.090	.044
PM Recognition	-.243	.006	.223	-.220	-.325	-.030	.310	-.077	.171
VR I	**-.357**	**.478**	**.387**	.023	.329	.298	-.301	.341	-.054
VR II	**-.422**	.258	**.381**	.257	-.141	-.073	**-.588**	-.044	.409
BVMT-R Immediate	**-.493**	**.420**	**.461**	-.176	.400	.416	**-.495**	.039	-.032
BVMT-R Delayed	**-.483**	**.369**	**.379**	-.097	.316	**.512**	**-.466**	.064	-.274

Note. Ed = years of education; DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, 4th Edition (WMS-IV) Visual Reproduction I; VR II = WMS-IV Visual Reproduction II; BVMT-R = Brief Visuospatial Memory Test, Revised Edition.

Recognition measures. Partial correlations were also calculated for raw scores on Visual Reproduction Recognition and BVMT-R Recognition Hits. The two subtests were moderately and positively correlated to each other in the clinical group (r =.335), although this result was not

Table 19

Partial Intercorrelations of Scaled Score Performance for the Control and Clinical Groups

	Subtest							
Subtest	1	2	3	4	5	6	7	8
1. WRAML2 Design Memory		***.608***	**.761**	-.397	**.622**	.337	**.750**	**.512**
2. WRAML2 Picture Memory	-.180		**.561**	.063	**.589**	.251	**.509**	.256
3. WRAML2 Design Memory Recognition	.263	.093		-.127	.308	.097	***.740***	.363
4. WRAML2 Picture Memory Recognition	.023	.431	-.041		.091	-.194	.006	-.056
5. WMS-IV Visual Reproduction I	***.732***	-.249	.350	.159		.343	**.677**	**.681**
6. WMS-IV Visual Reproduction II	.164	.279	.020	.204	.130		.139	.095
7. BVMT-R Immediate Recall	***.785***	-.122	.021	-.351	.435	.229		***.786***
8. BVMT-R Delayed Recall	**.687**	-.017	.330	-.318	.445	.314	***.851***	

Note. Unshaded = clinical group, shaded = control group; bold text indicates significance at or below the $p = .05$ level; bold and italic text indicates significance at or below the $p = 0.0018$ level. WRAML2 = Wide Range Assessment of Memory and Learning, 2nd Edition; WMS-IV = Wechsler Memory Scale, 4th Edition; BVMT-R = Brief Visuospatial Memory Test, Revised Edition.

statistically significant ($p = .189$). The two subtests were moderately and positively correlated to each other in the control group, as well ($r = .542$), and this result reached significance ($p = .025$).

Overall, measures are more strongly correlated in the control group than in the clinical group, even after controlling for the influence of age, education, and estimated premorbid IQ. It is not clear whether or why these differences exist, but it is possible that they may reflect differences in variability between the groups, i.e., the relative homogeneity of the control group

versus the difference in scores in the clinical group, in which participants presumably have differing levels of impairment and therefore a wider range of performance on each measure. Indeed, standard deviations for most measures are greater in the clinical group than in the control group.

Hypotheses 3 and 4: Comparisons of Mean Scaled Scores between Measures

It was hypothesized that different measures of visual memory are equivalent. Given the moderate to strong relationships between some of the visual memory measures (Tables 17 and 19), one way to demonstrate equivalence between measures is to assume that two measures are measuring the same construct when they are strongly correlated. A significant correlation between two measures would indicate that the two measures are measuring the same construct in a similar fashion. Correlation coefficients in bold text (with or without italics) in Tables 17 and 19 indicate that the two measures were found to be strongly correlated and therefore equivalent *within* each group in the present study. Because there were 28 correlations, however, a Bonferroni correction was necessary to reduce the likelihood of Type I error; consequently, using an using an α of .05 / 28 = .0018, those correlation coefficients that are bolded and italicized in Table 19 can be considered equivalent to each other within a similar group of patients/controls. Of note is that the strongest correlations were found between free-recall measures in the clinical group (WRAML2 Design Memory, WMS-IV Visual Reproduction I, and BVMT-R Immediate and Delayed Recall). An interesting finding is the strength of the correlation between BVMT-R Immediate and Delayed Recall scores in the clinical group. Either participants in the clinical group performed equally poorly on both measures, or the measures are sufficiently similar as to be redundant.

When making comparisons *between* groups, another way to demonstrate equivalence is to compare the strength of the correlation between two measures within each subgroup. If the correlation coefficient between two measures in one group does *not* differ significantly from the correlation coefficient between the same two measures in the other group, the measures may be considered equivalent in that group. A significant *difference* in the strength of the two correlations would therefore imply that two measures are *not* equivalent according to this criterion. To determine which measures were equivalent according to this criterion, partial correlations were converted to *z* scores using a Fisher transformation. The strength of the relationship between several pairs of measures differed significantly between groups (Table 20), suggesting that the pair of measures may not be considered equivalent, at least in this particular sample. Using a Bonferroni-corrected α of .0018, however, no differences between partial correlation coefficients were significant.

By this metric, large differences between correlation coefficients were found between groups for the following pairs of measures: WRAML2 Design Memory and Picture Memory; WRAML2 Design Memory and Design Memory Recognition; WRAML2 Picture Memory and WMS-IV Visual Reproduction I; WRAML2 Picture Memory and BMVT-R Immediate Recall; and WRAML2 Design Memory Recognition and BVMT-R Immediate Recall. Notably, all of these pairs of measures differ in terms of either item content, response format, or both; no significant differences between correlations were found for pairs of measures that are similar in content and structure (e.g., WRAML2 Design Memory and WMS-IV Visual Reproduction I).

When comparing normal to impaired groups, another way to determine equivalence between measures is to compare the degree to which the mean scaled scores on each measure

Table 20

Comparison of Partial Correlation Coefficients of Clinical and Control Groups

Measure 1	Measure 2	$r_{clinical}$	$r_{control}$	z	p
DM	PM	-.180	.608	-2.59	**.005**
	DM Recognition	.263	.761	-2.13	**.017**
	PM Recognition	.023	-.397	1.29	.099
	VR I	.732	.622	0.60	.275
	VR II	.164	.337	-0.54	.295
	BVMT-R Immediate	.785	.750	0.25	.402
	BVMT-R Delayed	.687	.512	0.81	.210
PM	DM Recognition	.093	.561	-1.58	.057
	PM Recognition	.431	.063	1.16	.123
	VR I	-.249	.589	-2.71	**.003**
	VR II	.279	.251	0.09	.466
	BVMT-R Immediate	-.122	.509	-1.99	**.023**
	BVMT-R Delayed	-.017	.256	-0.81	.208
DM Recognition	PM Recognition	-.041	-.127	0.25	.400
	VR I	.350	.308	0.14	.446
	VR II	.020	.097	-0.23	.411
	BVMT-R Immediate	.021	.740	-2.71	**.003**
	BVMT-R Delayed	.330	.363	-0.11	.457
PM Recognition	VR I	.159	.091	0.20	.420

Table 20 Continued

Measure 1	Measure 2	$r_{clinical}$	$r_{control}$	z	p
	VR II	.204	-.194	1.18	.120
	BVMT-R				
	Immediate	-.351	.006	-1.09	.139
	BVMT-R				
	Delayed	-.318	-.056	-0.80	.213
VR I	VR II	.130	.343	-0.66	.254
	BVMT-R				
	Immediate	.435	.677	-1.04	.149
	BVMT-R				
	Delayed	.445	.671	-0.97	.165
	BVMT-R				
VR II	Immediate	.229	.139	0.27	.393
	BVMT-R				
	Delayed	.314	.095	0.67	.251
	BVMT-R				
BVMT-R Immediate	Delayed	.851	.786	0.58	.281

Note. Bold text represents correlation coefficients that are significant at or below the $p = .05$ level. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, Revised Edition (WMS-R) Visual Reproduction I; VR II = WMS-R Visual Reproduction II.

differ between groups. By definition, patients with AD exhibit some degree of memory impairment. Participants in this study were grouped according to the presence or absence of a diagnosis of AD. The clinical group was expected to perform below the level of the control group, although it was not known to what degree they would underperform. For practical

purposes, any measure whose mean scaled scores differ significantly between groups of memory-impaired and presumably memory-unimpaired participants can be considered equivalent to any other measure that meets such a criterion. In other words, if the clinical goal is to document impairment in memory as defined by performance on a measure that deviates significantly from a population mean, any measure that reliably does so can be used. In order to make these comparisons, independent-samples *t*-tests were conducted. Using a Bonferroni-corrected α of .05 / 8 = .006, mean scaled scores differed significantly between groups on all measures for which scaled scores were derived ($p < .001$)

Levene's Test for Equality of Variances was significant for Picture Memory Recognition, ($p = .021$), and for BVMT-R Delayed ($p = .016$). Therefore, Mann-Whitney independent-samples *U*-tests were conducted. Mean scaled scores on Picture Memory Recognition were significantly greater in the control group than in the clinical group ($U(38) = 67.5$, $z = -3.619$, $p < .001$), and mean scaled scores on BVMT-R Delayed were significantly greater in the control group than in the clinical group ($U(38) = 44.0$, $z = -4.226$, $p < .001$). By this metric, therefore, each measure demonstrates reliable differences between normal and impaired groups.

Yet another way to demonstrate equivalence between measures is to show that mean scaled scores for different measures differ equally *between* groups but not *within* groups. In other words, if mean scaled scores do not differ significantly between measures within a particular group, measures can be considered equivalent in the sense that each one will yield the same conclusion when used alone to determine the relative strength or weakness of a participant's performance relative to a control group of normal healthy volunteers (as in the standardization samples for each measure).

A 2 × 8 mixed within-between groups repeated-measures analysis of variance was conducted in order to determine whether there were meaningful differences between measures in the overall sample and within each subgroup. This method compares the differences in mean scaled scores between measures to each other, both in the entire sample and within or between each subgroup, thereby determining whether any of these differences are statistically significant. The assumption of equal dependent variable covariance matrices across groups was met (Box's $M = 76.101$, $F(36,4859) = 1.632$, $p = .010$). The assumption of sphericity was violated (Mauchly's $W = .094$, $p < .001$). Therefore, multivariate tests were used in the analysis. There was a significant within-subjects main effect for measure (Wilks' $\lambda = .308$, $F(7,32) = 10.265$, $p < .001$, partial $\eta^2 = .692$), with a power to detect significance of 1.000. There was a significant between-subjects main effect ($F(1) = 84.644$, $p < .001$, partial $\eta^2 = .690$), with a power to detect significance of 1.000. There was a significant within-between interaction effect (Wilks' $\lambda = .545$, $F(7,32) = 3.822$, $p = .004$, partial $\eta^2 = .455$), with a power to detect significance of .948. This interaction suggests that differences observed between mean scaled scores on visual memory measures between groups do vary (as can be seen in Figure 1); however, since paired comparisons show each test's result is significantly different between groups, the significant main effect of group can be interpreted. Results of the ANOVA indicate that the differences observed between mean scaled scores on visual memory measures are strongly influenced by group membership (i.e., diagnosis). While there is a difference in level of performance between measures, the significant interaction effect limits the interpretability of this finding.

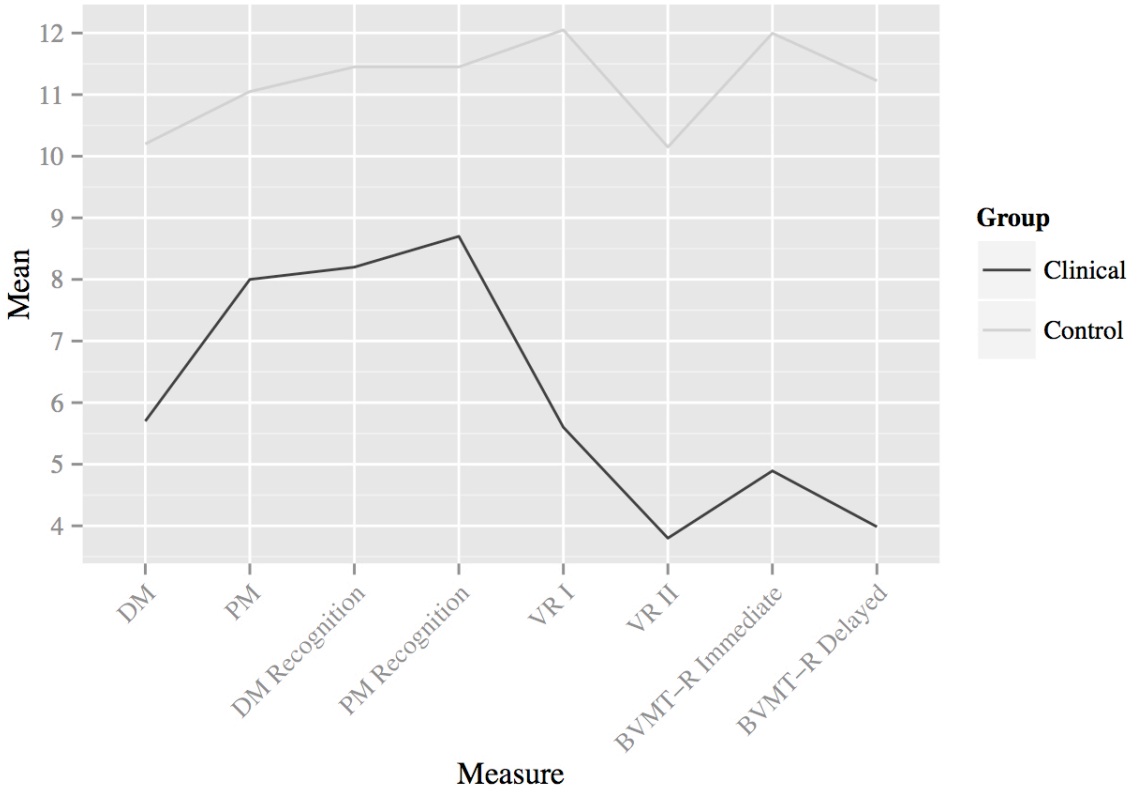

Figure 1. Mean Scaled Scores on Visual Memory Measures by Group.

Note. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, 4th Edition (WMS-IV) Visual Reproduction I; VR II = WMS-IV Visual Reproduction II; BVMT-R = Brief Visuospatial Memory Test, Revised Edition.

Because of the moderate to strong correlations that were found between demographic variables (age, education, estimated premorbid IQ) and the visual memory measures, and because those demographic variables differed significantly between groups, a 2 × 8 mixed within-between groups repeated-measures analysis of covariance controlling for age, education, and premorbid IQ was conducted in order to determine whether an interaction effect or any main effects would remain after controlling for those variables. The assumption of equal dependent

variable covariance matrices across groups was met (Box's $M = 76.101$, $F(36,4859) = 1.632$, $p = .010$). The assumption of sphericity was violated (Mauchly's $W = .094$, $p < .001$). Therefore, multivariate tests were used in the analysis. There was no significant within-subjects main effect for measure (Wilks' $\lambda = .843$, $F(7,29) = 0.774$, $p = .614$, partial $\eta^2 = .157$). However, the within-subjects analysis yielded a power to detect significance of only .272 at an α level of .05, indicating that the sample size was not sufficiently large to detect a difference between measures after removing covariates. There was a significant between-subjects main effect ($F(1) = 50.364$, $p < .001$, partial $\eta^2 = .590$), with a power to detect significance of 1.000. There was a significant within-between interaction effect (Wilks' $\lambda = .630$, $F(7,29) = 2.435$, $p = .043$, partial $\eta^2 = .370$), with an observed power to detect significance of .771 at an α level of .05, suggesting that differences observed between mean scaled scores on visual memory measures are strongly influenced by group membership, even after removing covariates. It appears likely that group membership is the source of the differences between mean scaled scores on each measure when measures are compared within the entire sample. Pair-wise comparisons indicate that some of the largest between-groups differences were found between WMS-IV Visual Reproduction II and WRAML2 subtests, as shown in Table 21.

As previously mentioned, the significant interaction effect observed suggests that mean scaled scores differ between visual memory measures to varying degrees depending on group membership. In order to identify the specific measures that contributed to this effect, 1 × 8 repeated-measures ANOVAs were conducted for each group. For the clinical group, the assumption of sphericity was violated (Mauchly's $W = .018$, $p < .001$). Therefore, multivariate tests were used in the analysis. There was a significant main effect for measure (Wilks' $\lambda = .142$,

Table 21

Pair-wise Comparisons between Measures in the Overall Sample Controlling for Age, Education, and Estimated Premorbid IQ

Measure 1	Measure 2	Mean Difference	*p*
DM	PM	-1.575	.081
	DM Recognition	-1.875	**.001**
	PM Recognition	-2.125	**.012**
	VR I	-0.875	1.00
	VR II	0.975	.749
	BVMT-R Immediate	-0.495	1.000
	BVMT-R Delayed	0.344	1.000
PM	DM Recognition	-0.300	1.000
	PM Recognition	-0.550	1.000
	VR I	0.700	1.000
	VR II	2.550	**< .001**
	BVMT-R Immediate	1.081	1.000
	BVMT-R Delayed	1.919	.110
DM Recognition	PM Recognition	-0.250	1.000
	VR I	1.000	1.000
	VR II	2.850	**< .001**
	BVMT-R Immediate	1.381	.537
	BVMT-R Delayed	2.219	**.029**
PM Recognition	VR I	1.250	1.000
	VR II	3.100	**< .001**
	BVMT-R Immediate	1.631	.970
	BVMT-R Delayed	2.469	.066
VR I	VR II	1.850	.153
	BVMT-R Immediate	0.381	1.000

Table 21 Continued

Measure 1	Measure 2	Mean Difference	*p*
	BVMT-R Delayed	1.219	.494
VR II	BVMT-R Immediate	-1.470	.486
	BVMT-R Delayed	-0.631	1.000
BVMT-R Immediate	BVMT-R Delayed	0.839	.668

Note. Bold text represents correlation coefficients that are significant at or below the $p = .05$ level after using a Bonferroni correction for multiple comparisons. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, Revised Edition (WMS-R) Visual Reproduction I; VR II = WMS-R Visual Reproduction II.

$F(7,13) = 11.182$, $p < .001$, partial $\eta^2 = .858$), with a power to detect significance of 1.00. For the control group, the assumption of sphericity was violated (Mauchly's $W = .023$, $p < .001$). Therefore, multivariate tests were used in the analysis. The main effect for measure was not significant (Wilks' $\lambda = .444$, $F(7,13) = 2.321$, $p = .09$, partial $\eta^2 = .556$), with a power to detect significance of .615. Pair-wise comparisons are shown in Table 22.

In the clinical group, the mean scaled score for the BVMT-R Delayed Recall was significantly lower than the mean scaled score for all four of the WRAML2 measures, and the mean scaled score for WMS-IV Visual Reproduction II was significantly lower than the mean scaled score for WRAML2 Design Memory Recognition, Picture Memory, and Picture Memory Recognition. In the control group, no mean scaled scores differed significantly. Pair-wise comparisons are shown in Table 23.

The large difference in scores between WRAML2 subtests and the other two delayed recall subtests suggests that the WRAML2 may exhibit better measurement characteristics at

Table 22

Pairwise Comparisons between Measures in the Control Group

Measure 1	Measure 2	Mean Difference	*p*
DM	PM	-0.850	1.000
	DM Recognition	-1.250	.523
	PM Recognition	-1.250	1.000
	VR I	-1.850	1.000
	VR II	0.050	1.000
	BVMT-R Immediate	-1.797	.537
	BVMT-R Delayed	-1.027	1.000
PM	DM Recognition	-0.400	1.000
	PM Recognition	-0.400	1.000
	VR I	-1.000	1.000
	VR II	0.900	1.000
	BVMT-R Immediate	-0.947	1.000
	BVMT-R Delayed	-0.177	1.000
DM Recognition	PM Recognition	0.000	1.000
	VR I	-0.600	1.000
	VR II	1.300	1.000
	BVMT-R Immediate	-0.547	1.000
	BVMT-R Delayed	0.223	1.000
PM Recognition	VR I	-0.600	1.000
	VR II	1.300	1.000
	BVMT-R Immediate	-0.547	1.000
	BVMT-R Delayed	0.223	1.000
VR I	VR II	1.900	1.000
	BVMT-R Immediate	0.054	1.000
	BVMT-R Delayed	0.823	1.000

Table 22 Continued

Measure 1	Measure 2	Mean Difference	*p*
VR II	BVMT-R Immediate	-1.847	.858
	BVMT-R Delayed	-1.077	1.000
BVMT-R Immediate	BVMT-R Delayed	0.770	1.000

Note. Bold text represents correlation coefficients that are significant at or below the $p = .05$ level after using a Bonferroni correction for multiple comparisons. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, Revised Edition (WMS-R) Visual Reproduction I; VR II = WMS-R Visual Reproduction II.

lower levels of performance. In order to determine whether differences exist between the AD and MCI subgroups within the clinical group (which, presumably, should perform somewhat differently due to differing levels of impairment), independent-samples *t*-tests were conducted using a Bonferroni correction for multiple comparisons. When comparing the 6 participants in the clinical group who were diagnosed with MCI to the 14 who were diagnosed with AD, no differences between means were found for any of the visual memory measures, most likely due to the small number of participants in each group. Effect sizes, while modest, suggest that, although WRAML2 Picture Memory and Picture Memory Recognition scaled scores were higher than scaled scores on the other measures *within* the clinical group, they better differentiated between the small AD and MCI subgroups within the clinical group.

A 2 × 8 mixed within-between repeated-measures analysis of covariance controlling for age, education, and estimated premorbid IQ was conducted *within* the clinical group to determine whether there were any differences in performance between those participants diagnosed with AD *versus* those diagnosed with MCI. The within-subjects main effect for measure approached

Table 23

Pairwise Comparisons between Measures in the Clinical Group

Measure 1	Measure 2	Mean Difference	*p*
DM	PM	-2.30	.342
	DM Recognition	-2.50	**.042**
	PM Recognition	-2.00	**.046**
	VR I	0.10	1.000
	VR II	1.90	.203
	BVMT-R Immediate	0.81	1.000
	BVMT-R Delayed	1.72	**.012**
PM	DM Recognition	-0.20	1.000
	PM Recognition	-0.70	1.000
	VR I	2.40	1.000
	VR II	4.20	**< .001**
	BVMT-R Immediate	3.11	.230
	BVMT-R Delayed	4.02	**.003**
DM Recognition	PM Recognition	-0.50	1.000
	VR I	2.60	.164
	VR II	4.40	**< .001**
	BVMT-R Immediate	3.31	.077
	BVMT-R Delayed	4.22	**< .001**
PM Recognition	VR I	3.10	.203
	VR II	4.90	**< .001**
	BVMT-R Immediate	3.81	.130
	BVMT-R Delayed	4.72	**.003**
VR I	VR II	1.80	1.000
	BVMT-R Immediate	0.71	1.000
	BVMT-R Delayed	1.62	.981

Table 23 Continued

Measure 1	Measure 2	Mean Difference	*p*
VR II	BVMT-R Immediate	-1.09	1.000
	BVMT-R Delayed	-0.19	1.000
BVMT-R Immediate	BVMT-R Delayed	0.91	1.000

Note. Bold text represents correlation coefficients that are significant at or below the $p = .05$ level after using a Bonferroni correction for multiple comparisons. DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, Revised Edition (WMS-R) Visual Reproduction I; VR II = WMS-R Visual Reproduction II.

significance (Wilks' $\lambda = .320$, $F(7,9) = 2.735$, $p = .081$, partial $\eta^2 = .680$), with a power to detect significance of .603. There was no significant between-subjects main effect ($F(1) = 0.332$, $p = .573$, partial $\eta^2 = .022$), with a power to detect significance of .084. There was no significant within-between interaction effect (Wilks' $\lambda = .619$, $F(7,9) = 0.790$, $p = .613$, partial $\eta^2 = .381$), with a power to detect significance of .190. Results are shown in Table 24.

Recognition measures. To compare those recognition tasks for which scaled scores could not be derived (WMS-IV and BVMT-R), receive-operator characteristic (ROC) curves were conducted on subtest raw scores. Briefly, ROC curves compare the diagnostic accuracy of a test in determining whether a participant belongs to a certain group (e.g., normal or impaired) according to a given cut score (Hanley & McNeil, 1983). The ROC curve analyses compare each measure's ability to classify participants into one or the other group; classification accuracy (sensitivity and specificity) can then be compared between measures. In most cases, a measure may be considered clinically useful if its sensitivity and specificity both exceed 80%, although there are reasons why different cut scores may be selected to minimize false positives or false

Table 24

Mean Scaled Score Differences between AD and MCI Patients in the Clinical Group

	AD			MCI			Mean		
Measure	Mean	SD	*N*	Mean	SD	*N*	difference	*p*	*d*
DM	5.57	2.344	14	6.00	2.098	6	0.429	.704	-.19
PM	8.36	2.620	14	7.17	2.639	6	-1.190	.365	.45
DM Recognition	8.00	2.828	14	8.67	2.066	6	0.667	.611	-.27
PM Recognition	8.21	2.359	14	9.83	3.764	6	1.619	.255	-.52
VR I	5.64	4.144	14	5.50	2.429	6	-0.143	.939	.04
VR II	4.07	1.900	14	3.17	2.483	6	-0.905	.384	.41
BVMT-R Immediate	4.69	3.640	14	5.37	3.898	6	0.682	.711	-.18
BVMT-R Delayed	3.86	2.682	14	4.29	2.247	6	0.429	.736	-.17

Note. AD = Alzheimer's disease; MCI = mild cognitive impairment; DM = Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML2) Design Memory; PM = WRAML2 Picture Memory; VR I = Wechsler Memory Scale, Revised Edition (WMS-R) Visual Reproduction I; VR II = WMS-R Visual Reproduction II.

negatives (Matthews & Farewell, 2007). Equivalence between recognition measures can be demonstrated if no significant differences in classification accuracy can be demonstrated.

Raw scores for WMS-IV Visual Reproduction Recognition, BVMT-R Hits, BVMT-R False Alarms, and BVMT-R Discrimination were compared using MedCalc. For Visual Reproduction Recognition, using a cut score of ≤ 4 correct (out of 7 multiple-choice items) yielded an area under the curve (AUC) of .988 (Youden's $J = .90$, $p < .001$, sensitivity = 90%,

specificity = 100%). For BVMT-R Hits, using a cut score of ≤ 4 (out of 6 yes-no items) yielded an AUC of .620 (Youden's $J = .20$, $p = .180$, sensitivity = 30%, specificity = 90%). For BVMT-R False Alarms, using a cut score of 0 (i.e., a score of ≥ 1 classifies a participant as normal) yielded an AUC of .831 (Youden's $J = .60$, $p < .001$, sensitivity = 80%, specificity = 80%). For BVMT-R Discrimination Index, using a cut score of ≤ 4 yielded an AUC of .840 (Youden's $J = .50$, $p < .001$, sensitivity = 70%, specificity = 80%). Tables 25 and 26 summarize the sensitivity and specificity characteristics of WMS-IV Visual Reproduction Recognition and BVMT-R Recognition False Alarms, which where the only two measures that yielded cut scores with clinically-adequate characteristics.

Table 25

Sensitivity and Specificity of WMS-IV Visual Reproduction Recognition Raw Scores

Cut Score	Sensitivity (%)	Specificity (%)
0	0	100
≤ 4	90	100
≤ 5	100	75
≤ 7	100	0

Note: WMS-IV = Wechsler Memory Scale, 4th Edition

Comparisons were made between AUCs in order to determine whether any recognition measure was significantly better at classifying participants. The AUC for Visual Reproduction

Table 26

Sensitivity and Specificity of BVMT-R Recognition False Alarms Raw Scores

Cut Score	Sensitivity (%)	Specificity (%)
0	100	0
≥ 1	80	80
≥ 2	45	95
≥ 3	25	100
≥ 5	0	100

Note: BVMT-R = Brief Visuospatial Memory Test, Revised Edition

Recognition was significantly greater than for BVMT-R Hits (difference = 0.368, $p < .001$), for BVMT-R False Alarms (difference = 0.156, $p = .020$), and for BVMT-R Discrimination Index (difference = 0.148, $p = .015$). The area under the curve for BVMT-R Hits was significantly smaller than for BVMT-R False Alarms (difference = -0.211, $p = .032$) and for BVMT-R Discrimination Index (difference = -0.220, $p < .001$). The AUCs did not differ significantly between BVMT-R False Alarms and BVMT-R Discrimination (difference = -0.009, $p = .892$).

Of the recognition measures, Visual Reproduction Recognition most accurately classified normal and impaired participants in this sample. Using a cut score of ≤ 4 out of 7 correct responses (raw score), 90% of participants in the clinical group were correctly classified as impaired, and 100% of participants in the control group were classified as normal. In addition, a cut score of 0 on BVMT-R False Alarms (i.e., there were no 'Yes' responses to "foil" items that were not present on the original stimulus card) accurately classified 80% of participants as impaired when they belonged to the clinical group, and 80% of participants as normal when they

belonged to the control group. This approaches the balance of sensitivity and specificity observed among well-validated screening measures (Nasreddine et al., 2005; Solomon et al., 2014), suggesting a potential real-world application of that cut score. If the recognition portion of the BVMT-R is administered, a single false positive suggests the presence of memory impairment, at least in this small sample.

Overall, results of the ROC curve analyses suggest that the recognition measures are *not* equivalent. WMS-IV Visual Reproduction Recognition and BVMT-R Recognition False Alarms outperformed the other recognition measures for which scaled scores could not be derived (comparisons were not made between these measures and the WRAML2 recognition measures, which were grouped with the other measures for which scaled scores were available).

The WRAML2 Picture Memory subtest is based on recognition of changes between drawings, rather than on recall of the drawings themselves. Because random guessing may influence performance on this subtest, the number of commission errors made on the task was compared between groups. An independent-samples *t*-test found no significant difference in commission errors between groups ($t(38) = -0.733$, $p = .468$, $d = .23$). A receiver-operator characteristic (ROC) curve analysis yielded an area under the curve (AUC) of .525 ($p = .79$). At a cut score of ≥ 5, the number of commission errors distinguished normal from impaired participants with a sensitivity of 35% and a specificity of 80% (Youden's $J = .15$). These results suggest that both groups made about the same number of commission errors, on average, and commission errors do not appear to have exerted a significant influence on performance on this task.

Chapter 4

Discussion

The purpose of the study was to determine whether there was clinical equivalence among several commonly used visual memory measures, using patients with Alzheimer's disease or mild cognitive impairment, and a comparison, non-clinical age-matched group. While there are many ways to define equivalence, in this study four specific criteria for equivalence were explored: (a) correlations between measures; (b) similarity of correlations between measures in two groups; (c) differences between mean scaled scores between groups; and (d) differences between differences in mean scaled scores between groups. Each hypothesis aimed to compare selected visual memory measures according to each criterion for equivalence.

Hypothesis 1 stated that visual memory measures would be found to be significantly correlated. In general, the results of this study are consistent with test norming studies that report moderate to strong relationships between various visual memory measures (Benedict, 1997; Sheslow & Adams, 2003; Wechsler, 2009). The average *r* in the control group was .40, as compared to the average *r* in the clinical group of .29 (this was computed by subtracting the *r* for the control group from the *r* for the clinical group and averaging the absolute value of all difference scores). The following pairs of measures were strongly and positively correlated in the control group: WRAML2 Design Memory and Picture Memory; WRAML2 Design Memory and Design Memory Recognition; WRAML2 Design Memory and WMS-IV Visual Reproduction I; WRAML2 Design Memory and BVMT-R Immediate Recall; WRAML2 Picture Memory and

Design Memory Recognition; WRAML2 Picture Memory and WMS-IV Visual Reproduction I; WRAML2 Picture Memory and BVMT-R Immediate Recall; WRAML2 Design Memory Recognition and BVMT-R Immediate Recall; WMS-IV Visual Reproduction I and BVMT-R Immediate Recall; and WMS-IV Visual Reproduction I and BVMT-R Delayed Recall (Table 19). Thus, all of the immediate recall measures were strongly and positively correlated in the control group, providing evidence of convergent validity in a group of normal controls. According to the first criterion, these measures can be considered equivalent in a group of normal controls. Clinically, this adds to the existing literature suggesting that the selected visual memory measures measure the same construct despite differences in item content and response format.

While correlations were generally not as strong in the clinical group, strong positive correlations were found between the following pairs of measures: WRAML2 Design Memory and WMS-IV Visual Reproduction I; WRAML2 Design Memory and BVMT-R Immediate Recall; WRAML2 Design Memory and BVMT-R Delayed Recall; and BVMT-R Immediate and Delayed Recall. Clinically, this suggests that WRAML2 Design Memory, WMS-IV Visual Reproduction I, and the BVMT-R are equivalent according to the first criterion. The use of any of those measures will yield very similar scores in a group of patients with MCI and early AD, and it would not necessarily be beneficial to use both. However, because of the much weaker correlations between each of those measures' delayed retrieval components (Table 19), and because delayed retrieval is critical aspect of memory evaluation (Sheslow & Adams, 2003; Budson & Solomon, 2011), the results of this analysis cannot be used to make decisions about whether to use any or all of the delayed retrieval trials.

Hypothesis 2 stated that the strength of the relationships between different memory measures do not differ between clinical and non-clinical groups. Equivalence between measures would be demonstrated by a lack of difference between correlation coefficients. The results of this study provide reasonable grounds to reject this hypothesis. Differences were found in the strength of the relationships between some pairs of measures between groups. The average of the between-groups differences in correlation coefficients was .28 (again computed by subtracting the *r* for the control group from the *r* for the clinical group and averaging the absolute value of all difference scores). More specifically, correlations between WRAML2 Design Memory and Picture Memory, between WRAML2 Design Memory and Picture Memory Recognition, between WRAML2 Picture Memory and WMS-IV Visual Reproduction I, between WRAML2 Picture Memory and BVMT-R Immediate Recall, and between WRAML2 Design Memory Recognition and BVMT-R Immediate Recall were stronger in the control group than in the clinical group.

When making decisions between measures in the clinical setting, the measures that are most strongly correlated are more likely to be considered "equivalent" or "interchangeable, but this depends on the strength of the correlation. Thus, while Hypothesis 2 was rejected, it is not clear whether the results of the comparison between correlations between groups are sufficiently useful from a clinical perspective. Given the restricted range of scores in the clinical group, the strength of the correlations between the 3 immediate recall measures (WRAML2 Design Memory, WMS-IV Visual Reproduction I, and BVMT-R Immediate Recall) provide ample evidence that these measures are "equivalent" in a group of memory-impaired patients. However,

the fact that the correlations are equally strong in the control group provides evidence for equivalence according to this criterion.

Hypothesis 3 stated that for each measure, mean scaled scores were expected to differ between the clinical group and the control group. The results of this study definitively support this hypothesis. Differences in mean scaled scores between the clinical group and the control group ranged from 2.8 to 7.2 scaled score points, with corresponding effect sizes ranging from 1.26 to 2.77 (Table 9). Differences between mean scaled scores in the clinical group and a population mean scaled score of 10 ranged from 1.3 scaled score points on WRAML2 Picture Memory Recognition to 6.2 on WMS-IV Visual Reproduction II, with an average difference of 3.89 (Table 11). In general, differences were larger on the free recall measures, and in particular on the delayed recall measures. Thus, with the exception of WRAML2 Picture Memory Recognition, the performance of the clinical group on each measure fell below 1.5 scaled score points from the mean, a difference that represents a commonly-accepted criterion for memory *impairment* (i.e., 1.5 standard deviations below the mean; Budson & Solomon, 2011). Clinically, these results suggests that WRAML2 Design Memory, WMS-IV Visual Reproduction (I and II), and BVMT-R (Immediate and Delayed Recall) are equivalent in terms of their ability to document and quantify memory weaknesses in patients with memory impairment. Measures that rely on recognition (WRAML2 Picture Memory and the two WRAML2 delayed recognition measures) should only be used as adjuncts to one of the three recall-based measures, and not as a primary indicator of memory impairment.

Hypothesis 4 stated that differences between measures are larger in the clinical group than in the control group. Given the homogeneity of the control group, it is not surprising that

few differences were found between mean scaled scores across measures. This is a group of participants whose performance on memory measures should be average or above average for their age group. Accordingly, differences between measures in the clinical group, which was less homogeneous, were larger.

Clinically, the differences in differences between measures within each group do not necessarily imply that one is "better" than another to distinguish between groups. Rather, these differences may represent unique characteristics of the sample as well as differences in difficulty and floor effects between measures. The clinical group in this case was relatively high-functioning, and the performance of participants in the clinical group was greater than the mean scaled score of 10 on several measures (specifically, WRAML2 Design Memory Recognition and Picture Memory Recognition, WMS-IV Visual Reproduction I, and BVMT-R Immediate Recall; Table 10).

While there is some evidence demonstrating that both groups of participants performed differently on different measures, the overall evidence suggests that differences in performance are greatest in the clinical group. In the clinical group, the largest differences between subtest mean scaled scores were found between WRAML2 recognition measures and the other free recall measures. Overall, participants in the clinical group performed better on the WRAML2 recognition measures than on WRAML2 Design Memory, WMS-IV Visual Reproduction II, and BVMT-R Delayed Recall. Participants performed worse on BVMT-R Delayed Recall than on WRAML2 Design Memory, WRAML2 Picture Memory, WRAML2 Design Memory Recognition, and WRAML2 Picture Memory Recognition. No differences in performance were found between WMS-IV Visual Reproduction (I and II) and the BVMT-R (Immediate and

Delayed Recall). The average difference between mean scaled scores on each measure in the clinical group was 1.6 scaled score points, ranging from .10 (WRAML2 Design Memory vs. WMS-IV Visual Reproduction I) to 4.9 (WRAML2 Picture Memory Recognition vs. WMS-IV Visual Reproduction II). The most interesting finding is the poorer overall performance of the clinical group on the BVMT-R, particularly the Delayed Recall trial, as compared to the other measures. On the other hand, pair-wise comparisons between measures in the control group did not reveal any significant differences in mean scaled scores on any of the measures for which scaled scores were derived. Clinically, these results suggest that while the BVMT-R is not equivalent to WRAML2 measures in the memory-impaired group, it is equivalent to WMS-IV Visual Reproduction. In this case, the WRAML2 measures, especially WRAML2 Design Memory (which is similar to WMS-IV Visual Reproduction and the BVMT-R in content and response format), are different enough from the other measures that they may be considered useful additions to a clinical battery when time permits.

Finally, WMS-IV and BVMT-R recognition measures were compared separately, and results indicate that the WMS-IV Visual Reproduction recognition trial may be useful on its own to classify normal from impaired participants. However, its use is limited by the fact that it requires the prior administration of the full WMS-IV Visual Reproduction subtest (I and II).

Conclusions/Implications

Taken overall, the results of this study support the use of WRAML2 Design Memory in addition to either WMS-IV Visual Reproduction (I and II, along with Visual Reproduction Recognition) or the BVMT-R (Immediate and Delayed, along with Recognition). Each of these measures can be used independently to diagnose MCI or AD and document memory weaknesses.

Participants in both groups performed relatively similarly on WMS-IV Visual Reproduction and on the BVMT-R. The results of this study do not provide any empirical reason to suggest that either one is "better" than the other; they are therefore "equivalent." On the other hand, WRAML2 measures, especially WRAML2 Design Memory, demonstrated unique measurement characteristics that may make it a useful additions to existing test batteries. Empirically, WRAML2 Design Memory provides additional variability in scores at lower levels of performance, which may make it more reliable across a wider range of patients who may be more severely impaired than participants in the present study. For example, Perri, Serra, Carlesimo, and Caltagirone (2007) found that the participants who converted to AD from amnestic MCI (aMCI) performed worse on visual memory measures than those who did not convert. Therefore, subtle differences in test scores may be useful in making prognostic judgments, and the lower floor of WRAML2 measures may make such decisions more accurate. Clinically, WRAML2 Design Memory was easier and faster to administer and score than WMS-IV Visual Reproduction despite the similarities in content between the two measures.

Certain measures were found to have different measurement characteristics than other similar measures in the overall sample. The differences emerged primarily in the clinical group, where mean scaled scores were higher for the WRAML2 delayed recognition measures than for the WMS-IV and BVMT-R delayed recall measures. This finding is not surprising, given the fact that the WRAML2 delayed recognition measures use a binary yes-no response format, increasing the influence of chance (and "lucky guesses") on a participant's total score. Results confirm the fact that recall tasks are generally more difficult than recognition tasks. All measures demonstrated large differences in mean scaled scores between the clinical group and the control

group and between the clinical group and normative samples, suggesting that any given measure can be useful on its own. The differences in the clinical group suggest that when impairment is present, WMS-IV and BVMT-R delayed recall measures may overemphasize the impairment. In addition, differences between WRAML2, WMS-IV, and BVMT-R delayed retrieval trials are likely influenced by the fact that the WRAML2 uses recognition rather than recall to measure delayed retrieval.

There may be a floor effect on the other measures (i.e., both clinical subgroups performed equally poorly due to item difficulty or limited range with very low scores), once again suggesting that the measurement characteristics of the WRAML2 measures may allow the clinician to make more refined conclusions about degree of pathology. Further evaluation of this finding is warranted, particularly because of the potential ecological utility of a real-world, contextual visual recognition task such as Picture Memory.

The clinical group in this study was reasonably similar to the clinical groups in normative studies, with the exception of level of education; the clinical group in this study was more highly educated. While limited information is available about the nature of those samples in terms of severity of impairment or stage of disease, the clinical group in this study was comprised almost entirely of patients with early-stage AD or MCI due to AD. It is possible that the performance of the clinical group on WRAML2 subtests, particularly WRAML2 Design Memory, demonstrates a larger measurement floor for those measures, allowing for greater variability in scores at lower levels of performance as the disease progresses. This finding may have been influenced by an inflationary impact that commission errors can have on WRAML2 Picture Memory scores, but the number and, therefore, the impact of commission errors was not evaluated in this study.

Thus, when using such measures to document strengths and weaknesses rather than simply to classify normal vs. impaired, WRAML2 subtests may provide a more sensitive measure of weakness.

Anecdotally, WRAML2 Picture Memory appeared to be more enjoyable for those participants who felt that they did poorly on other design recall tasks. Picture Memory does possess face validity that may be clinically useful, given additional support for ecological validity. Specifically, very few older adults are likely to be expected to recall abstract geometric designs on a regular basis, while almost all adults must repeatedly identify and recognize details and changes in their immediate visual environment. Picture Memory may better assess this important cognitive ability, although this will require further ecological validity studies.

Other Useful Findings

Mean scaled scores for BVMT-R Immediate Recall and Delayed Recall were strongly correlated in both groups. In terms of classifying normal from impaired, the Delayed trial provides useful information about retention and recall, which is an important aspect of memory evaluation, as differences between encoding, retention, and recall can aid in differential diagnosis of underlying pathology (Budson et al., 2006). The strong correlation between immediate and delayed recall on the BVMT-R is worthy of additional evaluation in future studies.

In terms of recognition tasks, the WMS-IV Visual Reproduction Recognition raw score was the best at differentiating normal from impaired. It is the only recognition task that used a multiple-choice response format. Compared to yes/no recognition tasks, this measure likely allows for a more normally distributed range of scores despite having fewer total items.

Unfortunately, this measure can only be administered as part of the entire Visual Reproduction subtest, so the advantages it provides are lost by the fact that a lengthy, difficult task with an inadequate floor precedes it. Regarding the finding that a single false-positive on the BVMT-R Recognition task is indicative of cognitive impairment, this may reflect inefficient response inhibition in individuals with mild cognitive impairment (Wylie, Ridderinkhof, Eckerle, & Manning, 2007).

Limitations

This study presents a number of limitations that may compromise generalizability of the results. In general, lack of statistical power to detect differences between measures within groups is likely. The study was powered to detect differences *between* groups, which tend to be greater than .75 standard deviations in most normative studies. Differences in performance between measures *within* each group were much smaller, and null hypothesis tests failed to detect significance in all but a few cases. With sufficient power, it may have been possible to detect a difference between measures. In that case, however, the differences would be small, and the resources to detect them would increase exponentially. From a practical standpoint, then, it can be cautiously concluded that any statistically significant differences between measures found with a larger sample may have negligible clinical significance.

There is evidence to suggest that variability in neuropsychological test scores changes with increasing dementia severity (Reckess, Varvaris, Gordon, & Schretlen, 2014). Demographic variables such as age, education, and IQ strongly affect decisions about “normal” and “impaired” (Leckliter, 1989), and SES can also have an impact on the initial clinical presentation of patients with memory complaints (Qian, Schweizer, & Fischer, 2013). If someone does well on verbal

memory but not on visual memory, or vice-versa, it may be due to differences in spatial skills (Heilbronner, 1992). Visuospatial skills are often impaired in early AD (Iachini et al., 2009). This was not considered in the analysis. Both the WRAML2 and the WMS-IV have optional copy trials to identify potential visual-motor deficits, but these trials were not used in the present study.

Executive functioning was not taken into consideration, but it stands to reason that it would have an effect, given the possible differences in executive demand between measures (Busch et al., 2005). In particular, performance on the BVMT-R may be more susceptible to weaknesses in executive functioning, as stimuli are presented all at once on a large grid, requiring more effort on the part of the examinee to plan and organize an encoding strategy. Likewise, the effect of processing speed was not taken into consideration, despite some evidence that differences in processing speed may influence performance on memory measures (Tam & Schmitter-Edgecombe, 2013). While none of the responses on any of the measures is time-limited, stimuli are presented for only a short time (5-10 seconds), and the amount of visual information presented varies from task to task; it is therefore possible that subtle processing speed deficits may impair the examinee's ability to rapidly and efficiently encode the varying amounts of visual information depending on which measure is chosen.

Age, education, and IQ consistently influence performance on a broad range of neuropsychological measures (Lezak et al., 2004). Due in part to the correlation between IQ and level of education, the mean estimated premorbid IQ for the control group was significantly greater than the population mean standard score of 100. Furthermore, because of a potential WTAR ceiling effect (Spinks et al., 2009), there was limited variability in estimated premorbid

IQ within the control group. Given the well-documented relationships that exist between level of education and performance on a variety of neuropsychological tests (Lezak et al., 2004), this homogeneity limits any conclusions that can be made from control group data or comparisons between the control group and the clinical group, limiting generalizability of these results to groups of patients who are generally younger and more highly educated.

Because the control group was selected from a convenience sample, control group and clinical group participants were not matched in terms of demographic covariates or US Census data. The results of this study can be generalized only to older adults with probable AD or MCI and who match the demographic characteristics of this sample. The clinical group in this study had a mean level of education that was slightly greater than the general population. All participants were Caucasian. The control group was selected by convenience and is very unlikely to be a valid representation of the typical person who presents for evaluation and does not receive a diagnosis of Alzheimer's disease.

Compared to normative studies, some correlations were stronger in the control group, while others were weaker. This may be due in part to the homogeneity of the control group as well as differences between the control group and normative samples in terms of demographic variables. For example, the variability in scores on WRAML2 decreases as performance decreases because of age (Sheslow & Adams, 2003), thereby inflating the relationship between each score.

Both WRAML2 visual memory subtests and their recognition tasks were used, allowing for Visual Memory and Visual Recognition index scores to be derived for the present study. However, the BVMT-R does not use a subtest/index scoring paradigm, and only one visual

memory subtest from the WMS-IV was used. Analyses were therefore conducted on subtest scaled scores, which are less reliable than index scores (Adams & Reynolds, 2009). Comprehensive memory tests such as the WRAML2 and the WMS-IV are typically interpreted "top down," first at the index level and then, if indicated, at the subtest level. The results of this study are therefore limited by the use of scores that may not be as reliable as index scores.

Youngjohn, Larrabee, and Crook (1993) gathered normative data for the Revised Benton Visual Retention Test (BVRT) and found strong negative correlations between age and BVRT scores and strong positive correlations between education and BVRT scores. Rönnlund, Nyberg, Backman, and Nilson (2005) note that cohort differences in educational attainment are the reason for discrepancies between cross-sectional and longitudinal studies assessing the relationship between cognition and age. Particularly in the last half-century, level of education has been increasing, and this may partly be reflected in the sample—particularly in the control group, in which the difference in age between the oldest and youngest participants was almost 30 years. As mentioned above, the measure used to estimate premorbid IQ in this study (the WTAR) has a low ceiling that limits accuracy in estimating higher levels of performance. Estimating premorbid IQ is difficult in patients with degenerative neurological diseases. While an estimate of premorbid IQ is usefully in clinical settings, the cross-sectional nature of this study may have allowed for the use of a more reliable measure of current IQ to obtain better data, at least in the control group.

It should be noted that the BVMT-R measures were included in the diagnostic process that defined group membership. This may be why mean BVMT-R scaled scores were observed

to be lower than any other scores in the clinical group. However, clinicians making the diagnoses were blinded to the results of WRAML2 and WMS-IV measures.

The WRAML2 delayed visual memory subtests are recognition-only in format. It is therefore possible that the higher scores observed with these recognition subtests were due to the fact that recognition scores are likely artificially inflated by the 50% chance of accuracy afforded by the binary-choice format used by the DM and PM Recognition subtests. Scaled scores correct for this factor to a degree, but because the floor exists at 50% correct, there is less variability possible in obtained low raw scores, and greater possibility of atypical results with smaller samples.

References

Adams, W., & Reynolds, C. R. (2009). *Essentials of WRAML2 and TOMAL–2 assessment*. New York, NY: Wiley.

Adams, W., & deBros, G. B. (2010, August). A comparison of visual memory in Chinese and US children. Presentation at the 118th Annual Convention of the American Psychological Association, San Diego, CA.

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., … Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, *7*(3), 270–279.

Albert, M. S., Moss, M. B., Tanzi, R., & Jones, K. (2001). Preclinical prediction of AD using neuropsychological tests. *Journal of the International Neuropsychological Society*, *7*(5), 631–639.

Alladi, S., Arnold, R., Mitchell, J., Nestor, P. J., & Hodges, J. R. (2006). Mild cognitive impairment: Applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychological Medicine*, *36*(4), 507–515.

Alzheimer's Association. (2014). 2014 Alzheimer's disease facts and figures. *Alzheimer's and Dementia*, *10*(2), e47–e92.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., text rev.)*. Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, *57*(12), 1060–1073.

American Psychological Association. (2012). Guidelines for the evaluation of dementia and age-related cognitive change. *American Psychologist*, *67*(1), 1–9.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (eds.), *The psychology of learning and motivation* (Vol. 2). New York, NY: Academic Press.

Bäckman, L., Jones, S., Berger, A.-K., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*, *19*(4), 520–531.

Benedict, R. H. B. (1997). *Brief Visuospatial Memory Test–Revised professional manual*. Lutz, FL: Psychological Assessment Resources.

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

Berchtold, N. C., & Cotman, C. W. (1998). Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-Roman period to the 1960s. *Neurobiology of Aging*, *19*(3), 173–189.

Blackwell, A. D., Sahakian, B. J., Vesey, R., Semple, J. M., Robbins, T. W., & Hodges, J. R. (2004). Detecting dementia: Novel neuropsychological markers of preclinical Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *17*, 42–48.

Bornstein, R. A., & Chelune, G. J. (1988). Factor structure of the Wechsler Memory Scale–Revised. *The Clinical Neuropsychologist*, *2*(2), 107–115. doi:10.1080/13854048808520093

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*(1), 15–24. doi:10.1016/0887-6177(89)90003-6

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*(3), 339–344.

Buccione, I., Perri, R., Carlesimo, G. A., Fadda, L., Serra, L., Scalmana, S., & Caltagirone, C. (2007). Cognitive and behavioural predictors of progression rates in Alzheimer's disease. *European Journal of Neurology*, *14*, 440–446. doi:10.1111/j.1468-1331.2006.01693.x

Budson, A. E., & Price, B. H. (2005). Memory dysfunction. *New England Journal of Medicine*, *352*(7), 692–699.

Budson, A. E., & Solomon, P. R. (2011). *Memory loss: A practical guide for clinicians*. Philadelphia, PA: Elsevier Saunders.

Budson, A. E., & Solomon, P. R. (2012). New diagnostic criteria for Alzheimer's disease and mild cognitive impairment for the practical neurologist. *Practical Neurology*, *12*, 88–96.

Budson, A. E., Wolk, D. A., Chong, H., & Waring, J. D. (2006). Episodic memory in Alzheimer's disease: Separating response bias from discrimination. *Neuropsychologia*, *44*(12), 2222–2232.

Burkhart, S. O. (2011). *Using short-term memory measures to assess long-term memory in early-stage Alzheimer's disease* (Unpublished doctoral dissertation). George Fox University, Newberg, OR.

Burton, D. B., Mittenberg, W., & Burton, C. A. (1993). Confirmatory factor analysis of the Wechsler Memory Scale–Revised standardization sample. *Archives of Clinical Neuropsychology*, *8*(6), 467–475. doi:10.1093/arclin/8.6.467

Busch, R. M., Booth, J. E., McBride, A., Vanderploeg, R. D., Curtiss, G., & Duchnick, J. J. (2005). Role of executive functioning in verbal and visual memory. *Neuropsychology*, *19*(2), 171–180.

Cherner, M., Suarez, P., Lazzaretto, D., Artiola i Fortuny, L., Mindt, M. R., Dawes, S., … the HNRC Group. (2007). Demographically corrected norms for the Brief Visuospatial Memory Test–Revised and Hopkins Verbal Learning Test–Revised in monolingual Spanish speakers from the U.S.–Mexico border region. *Archives of Clinical Neuropsychology*, *22*, 343–353. doi:10.1016/j.acn.2007.01.009

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

De Anna, Francesca, Felician, O., Barbeau, E., Mancini, J., Didic, M., & Ceccaldi, M. (2014). Cognitive changes in mild cognitive impairment patients with impaired visual recognition memory. *Neuropsychology*, *28*(1), 98–105.

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed

clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society*, *9*, 936–946. doi:10.10170S1355617703960139

Deluca, J., & Cicerone, K. D. (1991). Construct validity of tests of verbal and visual mental imagery. *The Clinical Neuropsychologist*, *5*(2), 176–182.

Derflinger, S., Sorg, C., Gaser, C., Myers, N., Arsic, M., Kurz, A., … Mühlau, M. (2011). Grey-matter atrophy in Alzheimer's disease is asymmetric but not lateralized. *Journal of Alzheimer's Disease*, *25*(2), 347–357.

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*(5), 731–739.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. (H. A. Ruger & C. E. Bussenius, trans.). New York, NY: Teacher's College, Columbia University.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. New York, NY: Oxford University Press.

Eslinger, P. J., Damasio, A. R., Benton, A. L., & Van Allen, M. (1985). Neuropsychologic detection of abnormal mental decline in older persons. *JAMA*, *253*(5), 670–674. doi:doi:10.1001/jama.1985.03350290076029

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*(2), 171–191.

Frise, N. R. (2009). *Adapting short-term memory measures to assess long-term memory* (Unpublished doctoral dissertation). George Fox University, Newberg, OR.

Gale, S. D., Baxter, L., Connor, D. J., Herring, A., & Comer, J. (2007). Sex differences on the Rey Auditory Verbal Learning Test and the Brief Visuospatial Memory Test–Revised in the elderly: Normative data in 172 participants. *Journal of Clinical and Experimental Neuropsychology*, *29*(5), 561–567.

Giambra, L. M., Arenberg, D., Kawas, C., Zonderman, A. B., & Costa, P. T. (1995). Adult life span changes in immediate visual memory and verbal intelligence. *Psychology and Aging*, *10*(1), 123–139.

Golden, Z., Bouvier, M., Selden, J., Mattis, K., Todd, M., & Golden, C. (2005). Differential performance of Alzheimer's and vascular dementia patients on a brief battery of neuropsychological tests. *International Journal of Neuroscience*, *115*, 1569–1577. doi:10.1080/00207450590957953

Greene, J. D. W., Baddeley, A. D., & Hodges, J. R. (1996). Analysis of the episodic memory deficit in early Alzheimer's disease: Evidence from the doors and people test. *Neuropsychologia*, *34*(6), 537–551.

Gregory, R. J. (2007). *Psychological testing: History, principles, and applications* (5th ed.). Boston, MA: Allyn and Bacon.

Grundman, M., Petersen, R. C., Ferris, S. H., Thomas, R. G., Aisen, P. S., Bennett, D. A., … Thal, L. J. (2004). Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of Neurology*, *61*(1), 59–66.

Gu, Q., Dillon, C. F., & Burt, V. L. (2010). *Prescription drug use continues to increase: U.S. prescription drug data for 2007–2008* (NCHS Data Brief No. 42). Hyattsville, MD: National Center for Health Statistics.

Gurczynski, J. (2009). *Brief Visuospatial Memory Test–Revised normative data for ages 80-89*. (Unpublished doctoral dissertation). Pacific University, Forest Grove, OR.

Hall, S. J. (2006). *Comparison between memory test results among persons with Alzheimer's disease*. (Unpublished doctoral dissertation). George Fox University, Newberg, OR.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*(3), 839–843.

Heilbronner, R. L. (1992). The search for a "pure" visual memory test: Pursuit of perfection? *The Clinical Neuropsychologist*, *6*(1), 105–112.

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

Hoelzle, J. B., Nelson, N. W., & Smith, C. A. (2011). Comparison of Wechsler Memory Scale–-Fourth Edition (WMS–IV) and Third Edition (WMS–III) dimensional structures: Improved ability to evaluate auditory and visual constructs. *Journal of Clinical and Experimental Neuropsychology*, *33*(3), 283–291. doi:10.1080/13803395.2010.511603

Holdnack, J. A., Zhou, X., Larrabee, G. J., Millis, S. R., & Salthouse, T. A. (2011). Confirmatory factor analysis of the WAIS-IV/WMS-IV. *Assessment*, *18*(2), 178–191. doi:10.1177/1073191110393106

Iachini, T., Iavarone, A., Senese, V. P., Ruotolo, F., & Ruggiero, G. (2009). Visuospatial memory in healthy elderly, AD and MCI: A review. *Current Aging Science*, *2*, 43–59.

Jacova, C., Kertesz, A., Blair, M., Fisk, J. D., & Feldman, H. H. (2007). Neuropsychological testing and assessment for dementia. *Alzheimer's and Dementia*, *3*, 299–317. doi:10.1016/j.jalz.2007.07.011

Jaynes, J. (1976). *The origin of consciousness in the breakdown of the bicameral mind*. Boston, MA: Mariner Books.

Julien, R. M. (2007). *A primer of drug action: A comprehensive guide to the actions, uses, and side effects of psychoactive drugs* (11th ed.). New York, NY: Worth.

Kane, K. D. (2012). *Visual memory: Validity of BVMT–R and relationships with depressive and anxious symptoms* (Unpublished doctoral dissertation). University of Colorado Springs, Colorado Springs, CO.

Kane, K. D., & Yochim, B. P. (2014). Construct validity and extended normative data for older adults for the Brief Visuospatial Memory Test, Revised. *American Journal of Alzheimer's Disease and Other Dementias*. doi:10.1177/1533317514524812

Kawas, C. H., Corrada, M. M., Brookmeyer, R., Morrison, A., Resnick, S. M., Zonderman, A. B., & Arenberg, D. (2003). Visual memory predicts Alzheimer's disease more than a decade before diagnosis. *Neurology*, *60*, 1089–1093.

Kral, V. A. (1962). Senescent forgetfulness: Benign and malignant. *Canadian Medical Association Journal*, *86*(6), 257–260.

Larrabee, G. J., & Curtiss, G. (1995). Construct validity of various verbal and visual memory tests. *Journal of Clinical and Experimental Neuropsychology*,*17*(4), 536-547.

Larrabee, G. J., Kane, R. L., & Schuck, J. R. (1983). Factor analysis of the WAIS and Wechsler Memory Scale: An analysis of the construct validity of the Wechsler Memory Scale. *Journal of Clinical Neuropsychology*, *5*(2), 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*(3), 239–250.

Larrabee, G. J., Levin, H. S., & High, W. M. (1986). Senescent forgetfulness: A quantitative study. *Developmental Neuropsychology*, *2*(4), 373–385.

Larrabee, G. J., Trahan, D. E., & Curtiss, G. (1992). Construct validity of the Continuous Visual Memory Test. *Archives of Clinical Neuropsychology*, *7*(5), 395–405.

Leckliter. (1989). The influence of age, education, IQ, gender, and alcohol abuse on Halstead-Reitan Neuropsychological Test Battery performance. *Journal of Clinical Psychology*, *45*(4), 484–512.

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. *The Clinical Neuropsychologist*, *5*(1), 83–88. doi:10.1080/13854049108401845

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*, *3*(3), 239–249. doi:10.1037/0894-4105.6.3.239

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

Matthews, D. E., & Farewell, V. T. (2007). *Using and understanding medical statistics* (4th ed.). Basel, Switzerland: Karger.

McCoy, K. J. M. (2004). *Understanding the transition from normal cognitive aging to mild cognitive impairment: Comparing the intraindividual variability in cognitive functioning* (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., … Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, *7*(3), 263–269.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS–ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

Millis, S. R., Malina, A. C., Bowers, D. A., & Ricker, J. H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale–III. *Journal of Clinical and Experimental Neuropsychology*, *21*(1), 87–93.

Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron*, *20*, 445–468.

Morris, J. C., Storandt, M., Miller, J. P., McKeel, D. W., Price, J. L., Rubin, E. H., & Berg, L. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology*, *58*(3), 397–405.

Moye, J. (1997). Nonverbal memory assessment with designs: Construct validity and clinical utility. *Neuropsychology Review*, *7*(4), 157–170.

Müller, U., Winter, P., & Graeber, M. B. (2013). A presenilin 1 mutation in the first case of Alzheimer's disease. *The Lancet Neurology*, *12*(2), 129–130.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699.

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*(5), 387–393. doi:10.1016/0887-6177(92)90151-C

Obermann, K. R., Morris, J. C., & Roe, C. M. (2013). Exploration of 100 commonly used drugs and supplements on cognition in older adults. *Alzheimer's and Dementia*, *9*(6), 724–732.

Pachana, N. A., Boone, K. B., Miller, B. L., Cummings, J. L., & Berman, N. (1996). Comparison of neuropsychological functioning in Alzheimer's disease and frontotemporal dementia. *Journal of the International Neuropsychological Society*, *2*, 505–510.

Perri, R., Serra, L., Carlesimo, G. A., & Caltagirone, C. (2007). Amnestic mild cognitive impairment: difference of memory profile in subjects who converted or did not convert to Alzheimer's disease. *Neuropsychology*, *21*(5), 549.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*(3), 303–308.

Phillips, W. A., & Christie, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology*, *29*, 117–133.

Price, L. R., Tulsky, D., Millis, S., & Weiss, L. (2002). Redefining the factor structure of the Wechsler Memory Scale–III: Confirmatory factor analysis with cross-validation. *Journal of Clinical and Experimental Neuropsychology*, *24*(5), 574–585.

Qian, W., Schweizer, T. A., & Fischer, C. E. (2013). Impact of socioeconomic status on initial clinical presentation to a memory disorders clinic. *International Psychogeriatrics*, *26*(4), 597–603.

Querfurth, H. W., & LaFerla, F. M. (2010). Alzheimer's disease. *New England Journal of Medicine*, *362*(4), 329–344.

Reckess, G. Z., Varvaris, M., Gordon, B., & Schretlen, D. J. (2014). Within-person distributions of neuropsychological test scores as a function of dementia severity. *Neuropsychology*, *28*(2), 254–260.

Riege, W. H., & Inman, V. (1981). Age differences in nonverbal memory tasks. *Journal of Gerontology*, *36*(1), 51–58.

Rogers, S. L., Farlow, M. R., Doody, R. S., Mohs, R., & Friedhoff, L. T. (1998). A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. *Neurology*, *50*, 136–45.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilson, L. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging*, *20*(1), 3–18. doi:10.1037/0882-7974.20.1.3

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*(5), 754–760.

Salthouse, T. A., Fristoe, N., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? *Neuropsychology*, *10*(2), 272–285.

Sewell, K. W., Downey, R. G., & Sinnett, E. R. (1988). Convergence and divergence of clinical memory tests. *Psychological Reports*, *62*, 291–297.

Sheslow, D., & Adams, W. (2003). *Wide Range Assessment of Memory and Learning* (2nd ed.). Lutz, FL: Psychological Assessment Resources.

Sinnett, E. R., & Holen, M. C. (1999). Assessment of memory functioning among an aging sample. *Psychological reports*, *84*(1), 339–350.

Skilbeck, C. E., & Woods, R. T. (1980). The factorial structure of the Wechsler Memory Scale: Samples of neurological and psychogeriatric patients. *Journal of Clinical and Experimental Neuropsychology*, *2*(4), 293–300.

Smith, G. E., & Bondi, M. W. (2013). *Mild cognitive impairment and dementia*. New York, NY: Oxford University Press.

Solomon, T. M., deBros, G. B., Budson, A. E., Mirkovic, N., Murphy, C. A., & Solomon, P. R. (2014). Correlational analysis of 5 commonly used measures of cognitive functioning and mental status: An update. *American Journal of Alzheimer's Disease and Other Dementias*.

Soscia, S. J., Kirby, J. E., Washicosky, K. J., Tucker, S. M., Ingelsson, M., Hyman, B., … Moir, R. D. (2010). The Alzheimer's disease-associated amyloid -protein is an antimicrobial peptide. *PLoS ONE*, *5*(3), e9505.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., … Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease:

Recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, *7*(3), 280–292.

Spinks, R., McKirgan, L. W., Arndt, S., Caspers, K., Yucuis, R., & Pfalzgraf, C. J. (2009). IQ estimate smackdown: Comparing IQ proxy measures to the WAIS–III. *Journal of the International Neuropsychological Society*, *15*(4), 590–596.

Stelzmann, R. A., Schnitzlein, H. N., & Murtagh, F. R. (1995). An English translation of Alzheimer's 1907 paper, "Über eine eigenartige Erkrankung der Hirnrinde". *Clinical Anatomy*, *8*, 429–431.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.

Swainson, R., Hodges, J. R., Galton, C. J., Semple, J., Michael, A., Dunn, B. D., … Sahakian, B. J. (2001). Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Dementia and Geriatric Cognitive Disorders*, *12*, 265–280.

Tam, J. W., & Schmitter-Edgecombe, M. (2013). The role of processing speed in the Brief Visuospatial Memory Test–Revised. *The Clinical Neuropsychologist*, *27*(6), 962–972.

Tierney, M. C., Yao, C., Kiss, A., & McDowell, I. (2005). Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. *Neurology*, *64*(11), 1853–1859.

Troyer, A. K., Murphy, K. J., Anderson, N. D., Hayman-Abello, B. A., Craik, F. I. M., & Moscovitch, M. (2008). Item and associative memory in amnestic mild cognitive impairment: Performance on standardized memory tests. *Neuropsychology*, *22*(1), 10–16.

Tulsky, D. S., & Price, L. R. (2003). The joint WAIS–III and WMS–III factor structure: Development and cross-validation of a six-factor model of cognitive functioning. *Psychological Assessment*, *15*(2), 149–162. doi:10.1037/1040-3590.15.2.149

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (eds.), *Organization of memory* (pp. 381–402). New York, NY: Academic Press.

Walhovd, K. B., Fjell, A. M., Reinvang, I., Lundervold, A., Dale, A. M., Eilertsen, D. E., … Fischl, B. (2005). Effects of age on volumes of cortex, white matter and subcortical structures. *Neurobiology of Aging*, *26*, 1261–1270. doi:10.1016/j.neurobiolaging.2005.05.020

Wang, B.-S., Wang, H., Wei, Z.-H., Song, Y.-Y., Zhang, L., & Chen, H.-Z. (2009). Efficacy and safety of natural acetylcholinesterase inhibitor huperzine A in the treatment of Alzheimer's disease: An updated meta-analysis. *Journal of Neural Transmission*, *116*(4), 457–465.

Wechsler, D. (2009). *Wechsler Memory Scale* (4th ed.). San Antonio, TX: Pearson.

Wylie, S. A., Ridderinkhof, K. R., Eckerle, M. K., & Manning, C. A. (2007). Inefficient response inhibition in individuals with mild cognitive impairment. *Neuropsychologia*, *45*, 1408–1419. doi:10.1016/j.neuropsychologia.2006.11.003

Youngjohn, J. R., Larrabee, G. J., & Crook, T. H. (1993). New adult age- and education-correction norms for the Benton Visual Retention Test. *The Clinical Neuropsychologist*, *7*(2), 155–160.

Appendix A

DSM-IV-TR Diagnostic Criteria for Dementia of the Alzheimer's Type (DAT)

A. The development of multiple cognitive deficits manifested by both:
    1. memory impairment (impaired ability to learn new information or to recall previously learned information) and (2) one (or more) of the following cognitive disturbances:
        a. aphasia (language disturbance)
        b. apraxia (impaired ability to carry out motor activities despite intact motor function)
        c. agnosia (failure to recognize or identify objects despite intact sensory function)
        d. disturbance in executive functioning (i.e., planning, organizing, sequencing, abstracting).
B. The cognitive deficits in criteria A1 and A2 each cause significant impairment in social or occupational functioning and represent a significant decline from a previous level of functioning.
C. The course is characterized by gradual onset and continuing cognitive decline.
D. The cognitive deficits in criteria A1 and A2 are not due to any of the following:
    1. other central nervous system conditions that cause progressive deficits in memory and cognition (e.g., cerebrovascular disease, Parkinson's disease,

Huntington's disease, subdural hematoma, normal-pressure hydrocephalus, brain tumor)

2. systemic conditions that are known to cause dementia (e.g., hypothyroidism, vitamin B or folic acid deficiency, niacin deficiency, hypercalcemia, neurosyphilis, HIV infection)
3. substance-induced conditions.

E. The deficits do not occur exclusively during the course of a delirium.

F. The disturbance is not better accounted for by another Axis I disorder (e.g., major depressive episode, schizophrenia).

Adapted from American Psychiatric Association (2000).

Appendix B

NIA-AA Diagnostic Guidelines for Probable Alzheimer's Disease

Dementia is diagnosed when there are cognitive or behavioral (neuropsychiatric) symptoms that:

1. Interfere with the ability to function at work or at usual activities; and
2. Represent a decline from previous levels of functioning and performing; and
3. Are not explained by delirium or major psychiatric disorder.

The cognitive or behavioral impairment involves a minimum of two of the following domains:

1. Impaired ability to acquire and remember new information
2. Impaired reasoning and handling of complex tasks, poor judgment
3. Impaired visuospatial abilities
4. Impaired language functions (speaking, reading, writing)
5. Changes in personality, behavior, or comportment

Probable AD dementia is diagnosed when the patient's presentation meets criteria for dementia and, in addition, has the following characteristics:

1. Insidious onset;
2. Clear-cut history of worsening of cognition by report or observation; and
3. The initial and most prominent cognitive deficits are evident on history and examination in one of the following categories:
    a. Amnestic presentation: impairment in learning and recall of recently learned

information and evidence of cognitive dysfunction in at least one other cognitive domain.

b. Nonamnestic presentations:
    i. Language presentation: The most prominent deficits are in word-finding, but deficits in other cognitive domains should be present.
    ii. Visuospatial presentation: The most prominent deficits are in spatial cognition, including object agnosia, impaired face recognition, simultanagnosia, and alexia. Deficits in other cognitive domains should be present.
    iii. Executive dysfunction: The most prominent deficits are impaired reasoning, judgment, and problem solving. Deficits in other cognitive domains should be present.

4. The diagnosis of probable AD dementia should not be applied when there is evidence of
    a. substantial concomitant cerebrovascular disease, defined by
        i. a history of a stroke temporally related to the onset or worsening of cognitive impairment; or
        ii. the presence of multiple or extensive infarcts or severe white matter hyperintensity burden; or
    b. core features of Dementia with Lewy bodies other than dementia itself; or
    c. prominent features of behavioral variant frontotemporal dementia; or
    d. prominent features of semantic variant primary progressive aphasia or non-fluent/agrammatic variant primary progressive aphasia; or

e. evidence for another concurrent, active neurological disease, or a non-neurological medical comorbidity or use of medication that could have a substantial effect on cognition.

Adapted from McKhann et al. (2011).

Appendix C

DSM-5 Diagnostic Criteria for Major and Mild Neurocognitive Disorders

Major Neurocognitive Disorder

A. Evidence of significant cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition) based on:
    1. Concern of the individual, a knowledgeable informant, or the clinician that there has been a significant decline in cognitive function; and
    2. A substantial impairment in cognitive performance, preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment.
B. The cognitive deficits interfere with independence in everyday activities (i.e., at a minimum, requiring assistance with complex instrumental activities of daily living such as paying bills or managing medications).
C. The cognitive deficits do not occur exclusively in the context of a delirium.
D. The cognitive deficits are not better explained by another mental disorder (e.g., major depressive disorder, schizophrenia).

Minor Neurocognitive Disorder

A. Evidence of modest cognitive decline from a previous level of performance in one or more cognitive domains (complex attention, executive function, learning and memory, language, perceptual-motor, or social cognition) based on:

1. Concern of the individual, a knowledgeable informant, or the clinician that there has been a mild decline in cognitive function; and
2. A modest impairment in cognitive performance, preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment.

B. The cognitive deficits do not interfere with capacity for independence in everyday activities (i.e., complex instrumental activities of daily living such as paying bills or managing medications are preserved, but greater effort, compensatory strategies, or accommodation may be required).

C. The cognitive deficits do not occur exclusively in the context of a delirium.

D. The cognitive deficits are not better explained by another mental disorder (e.g., major depressive disorder, schizophrenia).

Major or Mild Neurocognitive Disorder Due to Alzheimer's Disease

A. The criteria are met for major or mild neurocognitive disorder.

B. There is insidious onset and gradual progression of impairment in one or more cognitive domains (for major neurocognitive disorder, at least two domains must be impaired).

C. Criteria are met for either probable or possible Alzheimer's disease as follows:

For major neurocognitive disorder: Probable Alzheimer's disease is diagnosed if either of the following is present; otherwise, possible Alzheimer's disease should be diagnosed.

1. Evidence of a causative Alzheimer's disease genetic mutation from family history or genetic testing.
2. All three of the following are present:
    a. Clear evidence of decline in memory and learning and at least one other cognitive domain (based on detailed history or serial neuropsychological testing).
    b. Steadily progressive, gradual decline in cognition, without extended plateaus.
    c. No evidence of mixed etiology (i.e., absence of other neurodegenerative or cerebrovascular disease, or another neurological, mental, or systemic disease or condition likely contributing to cognitive decline).

For mild neurocognitive disorder: Probable Alzheimer's disease is diagnosed if there is evidence of a causative Alzheimer's disease genetic mutation from either genetic testing or family history. Possible Alzheimer's disease is diagnosed if there is no evidence of a causative Alzheimer's disease genetic mutation from either genetic testing or family history, and all three of the following are present:

1. Clear evidence of decline in memory and learning.
2. Steadily progressive, gradual decline in cognition, without extended plateaus.
3. No evidence of mixed etiology (i.e., absence of other neurodegenerative or cerebrovascular disease, or another neurological or systemic disease or condition likely contributing to cognitive decline).

The disturbance is not better explained by cerebrovascular disease, another neurodegenerative disease, the effects of a substance, or another mental, neurological, or systemic disorder.

Adapted from American Psychiatric Association (2013).

Appendix D

Informed Consent to Participate in Clinical Research

Principal Investigator: Guy B. deBros, MA

The Memory Clinic

357 Shields Drive

Bennington, VT 05201

(802) 447-1409 x34

guy@memorydoc.org

I, _____, hereby consent to participate in The Memory Clinic's research on visual memory. I understand that testing will be conducted in conjunction with my diagnostic evaluation. I understand that the results of my evaluation will be used as part of the study. I understand that no identifiable information will be available within the study. I understand that all analyses will be conducted on aggregate data. I understand that there is minimal risk involved (i.e., fatigue due to an additional 5 to 10 minutes of testing). I understand that I may withdraw my consent to participate at any time and for any reason without any risk of consequences and without fear of jeopardizing my clinical relationship with The Memory Clinic. I understand that if I choose to withdraw my consent, my data will be removed from any and all analyses being conducted for the purpose of the study; however, I understand that withdrawal of consent to participate does not mean that records related to my clinical relationship with the

Memory Clinic will be destroyed. I understand that I may contact The Memory Clinic if I have any questions or concerns.

_____	_____
Printed Name of Participant	Signature of Participant or Legally Authorized Representative
_____	_____
Printed Name of Person Conducting Informed Consent Discussion (or P.I.)	Printed Name of Person Conducting Informed Consent Discussion (or P.I.)

Appendix E

Curriculum Vitae

Guy B. deBros
The Memory Clinic
357 Shields Drive
Bennington, VT 05201
(802) 447-1409
guy@memorydoc.org

Education

2009 – Present **George Fox University, Newberg, OR**
Graduate Dept. of Clinical Psychology: APA Accredited PsyD Program
Anticipated date of PsyD degree: November, 2014

2007 – 2009 **George Fox University, Newberg, OR**
Graduate Dept. of Clinical Psychology: APA Accredited PsyD Program
Master of Arts, Clinical Psychology

2001 – 2006 **Tufts University, Medford, MA**
Bachelor of Arts, Psychology/Clinical

Clinical Experience

2012 – Present **Clinical Psychology Doctoral Internship**
The Memory Clinic, Bennington, VT
Administered, scored, and interpreted neuropsychological tests. Participated in patient and caregiver interviews, diagnostic discussions, and case conferences. Prepared draft reports. Provided caregiver supportive counseling. Assisted with data entry and database management.
Supervisors: Cynthia Murphy, PsyD, MBA, Diana E. Michalczuk, PsyD, and Paul R. Solomon, PhD

2011 – 2012 **Clinical Psychology Practicum**
Providence Medical Group, Sherwood, OR
Provided behavioral health consultation services, brief cognitive-behavioral and solution-focused interventions, and psychoeducation in a suburban primary care clinic. Collaborated with staff to clarify diagnoses and make pharmacological and referral recommendations. Administered and interpreted

brief symptom checklists and screeners as needed.
Supervisor: Marie-Christine Rutter-Goodworth, PhD

2010 – 2012 **Clinical Psychology Practicum**
Sundstrom Clinical Services, Clackamas, OR
Provided psychoeducational and neuropsychological evaluation services in a child and adult outpatient group practice. Presenting problems included autism spectrum disorders, multiple sclerosis, ADHD, disruptive behavior disorders, depression and anxiety, as well as adjustment and personality disorders. Participated in weekly group didactic sessions. Weekly individual supervision with occasional group case discussions.
Supervisor: Paul E. Sundstrom, EdD

2009 – 2010 **Clinical Psychology Practicum**
Square Peg Psychological, Gladstone, OR
Provided school counseling services to female and male secondary students in a private high school setting. Provided consultation services to school staff. Presented didactic seminars to health and psychology classes. Co-facilitated a weekly skill-building group dealing with stress and anxiety. Weekly individual and group supervision.
Supervisor: Denise López Haugen, PsyD

2009 – 2010 **Clinical Psychology Practicum**
George Fox University Behavioral Health Clinic, Newberg, OR
Provided several dementia screening assessments and reports for patients referred by a psychiatrist at a local hospital. Presenting problems included closed-head injury, CVA, and multiple sclerosis. Supervision received for each referral.
Supervisor: Joel Gregor, PsyD

2009 – 2010 **Clinical Psychology Practicum**
St. Paul School District, St. Paul, OR
Provided several comprehensive psychoeducational assessments and reports at the high school and kindergarten levels. Participated in IEP meetings with school staff and family members. Provided feedback to school staff regarding possible interventions. Supervision received for each referral.
Supervisor: Elizabeth Hamilton, PhD

2008 – 2009 **Clinical Psychology Practicum**
Multnomah County Department of Corrections, Portland, OR
Provided psychological services to female and male adult inmates housed in a medium-security county jail. Direct services included individual psychotherapy (short- and long-term), psychological evaluations using personality, cognitive-intellectual, and neuropsychological measures,

consultations with medical staff, and recommendations for treatment placement. Co-facilitated weekly didactic group psychotherapy focusing on diverse topics including stress, anger management, relaxation, thinking skills, and assertiveness training. Presenting problems included multiple Axis I and II disorders and severe mental illness. Weekly individual and group supervision.
Supervisor: Stephen Huggins, PsyD, CCHP

2007 – 2008 **Clinical Psychology Pre-Practicum**
George Fox University, Newberg, OR
Provided outpatient psychotherapy services to female and male undergraduate students. Responsibilities included intake interviews, personality assessment, diagnosis, treatment planning, and individual psychotherapy. Monitored progress through video tape reviews and presented cases to supervision group.
Supervisors: Mary Peterson, PhD, Licensed Clinical Psychologist and Associate Director of Clinical Training; Lisa A. Jones, MA, Clinical Psychology Graduate Student

2006 – 2007 **Assistant Crisis Counselor**
S.E.E.K. Program, John F. Kennedy Elementary School, Somerville, MA
Paid position providing behavioral interventions to female and male special education students in a public elementary school. Responsibilities included assisting students with behavioral goals, providing support and intervening during behavioral crises, participating in weekly staff meetings, and supervising classroom activities alongside school staff.

2005 – 2006 **Clinical Psychology Undergraduate Practicum**
S.E.E.K. Program, John F. Kennedy Elementary School, Somerville, MA
Provided behavioral therapy services to female and male special education students in a public elementary school. Responsibilities included assisting students with behavioral goals, providing support and intervening during behavioral crises, participating in weekly staff meetings, and supervising classroom and extracurricular activities in conjunction with school staff. Weekly individual supervision.
Supervisor: Wanda Finigian, MEd, Licensed Learning Specialist

2005 – 2007 **Emergency Medical Technician – Basic**
EasCare Ambulance Service, Boston, MA
Provided emergency medical services and medical transportation to the Greater Boston area. Responsibilities included responding to life-threatening and non-life-threatening emergencies in a variety of settings and situations, monitoring basic vital signs before and during transportation to hospitals and other medical facilities, coordinating patient care with receiving providers,

and assisting paramedics and fire department workers during single- and multiple-casualty incidents.

2004 – 2005 **Emergency Medical Technician – Basic**
Professional Ambulance Service, Cambridge, MA
Provided emergency medical services and medical transportation to the City of Cambridge by ambulance. Responsibilities included responding to life-threatening and non-life-threatening emergencies in a variety of settings and situations, monitoring basic vital signs before and during transportation to hospitals and other medical facilities, coordinating patient care with receiving providers, and assisting paramedics and fire department workers during single- and multiple-casualty incidents.

2001 – 2004 **Anesthesia Technician**
Massachusetts General Hospital, Boston, MA
Provided supply and equipment support to anesthesiologists and nurse anesthetists in a suite of approximately 50 operating rooms. Responsibilities included cleaning and maintaining equipment, restocking supplies before and during operations, and assisting anesthesiologists and nurse anesthetists as needed during procedures.

Research and Professional Presentations

Solomon, T. M., **deBros, G. B.**, Budson, A. E., Mirkovic, N., Murphy, C. A., & Solomon, P. R. (2014). Correlational analysis of five commonly used measures of cognitive functioning and mental status: An update. *American Journal of Alzheimer's Disease and Other Dementias, 29*(8), 718-722.

deBros, G. B., & Solomon, T. M. (2014, February). The Montreal Cognitive Assessment (MoCA) is superior to the Mini-Mental State Examination (MMSE) in differentiating MCI from normal healthy aging in patients with subjective memory complaints. Presentation at the 42nd Annual Meeting of the International Neuropsychological Society, Seattle, WA.

Solomon, T. M., **deBros, G. B.**, Mirkovic, N., Murphy, C. A., & Solomon, P. R. (2014, February). Analysis of five commonly used measures of cognitive functioning and mental status. Presentation at the 42nd Annual Meeting of the International Neuropsychological Society, Seattle, WA.

Solomon, T. M., Budson, A. E., **deBros, G. B.**, Murphy, C. A., & Solomon, P. R. (2014, February). A proof of concept study for a randomized, double-blind, placebo-controlled, parallel group, efficacy study of Alpha BRAIN™ administered orally. Presentation at the 42nd Annual Meeting of the International Neuropsychological Society, Seattle, WA.

Solomon, T. M., **deBros, G. B.**, Mirkovic, N., Murphy, C., & Solomon, P. R. (2013, April). Relationships between commonly used measures of cognitive and functional status. Presentation at the Stanley Cobb Assembly of the Boston Society of Neurology and Psychiatry, Boston, MA.

Adams, W., & **deBros, G. B.** (2010, August). A comparison of visual memory in Chinese and US children. Presentation at the 118th Annual Convention of the American Psychological Association, San Diego, CA.

deBros, G. B., Jurecska, D. E., Millkey, A. M., & Peterson, M. (2010, May). The Malingered Ignorance of Legal Knowledge Test (MILK): A brief measure of forensic symptom validity. Presentation at the 10th Annual Conference of the International Association of Forensic Mental Health Services, Vancouver, BC, Canada.

University Involvement

Fall, 2011 **Co-Lecturer**
George Fox University, Newberg, OR
General Psychology
Prepared, co-taught, and managed an introductory psychology course as part of the graduate-level Academic Careers course.
Supervisor: Kathleen Gathercoal, PhD

2010 – 2012 **Graduate Teaching Assistant**
Graduate Dept. of Clinical Psychology, George Fox University, Newberg, OR
Neuropsychological Assessment
Coordinated class materials, taught and demonstrated the use of several neuropsychological tests.
Supervisor: Wayne Adams, PhD, ABPP

Professional Affiliations

2009 – Present **International Neuropsychological Society**
Student Member

2009 – Present **National Academy of Neuropsychology**
Student Member

2008 – Present **American Psychological Association**
Graduate Student Affiliate