Summer 2018

Screening for Dementia: An examination of subscale relative importance

Emily F. Matusz
James Madison University

Screening for Dementia: An Examination of Subscale Relative Importance

Emily F. Matusz

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

August 2018

FACULTY COMMITTEE:

Committee Chair: Bernice Marcopulos, Ph.D

Committee Members/ Readers:

Jeff Dyche, Ph.D

John Hathcoat, Ph.D

Scott Sperling, Psy.D

Acknowledgements

I would like to express my sincere appreciation to my advisor, Dr. Bernice Marcopulos, for her continuous support and dedication to student learning. Without her grounding guidance, I would likely still be stuck "spinning" in my thoughts while attempting to write page one of this project. Thank you, Bernice, for pushing me far beyond my own limits and for never accepting mediocrity. It has been such a privilege to work with and learn from you. I would also like to thank the members of my thesis committee, Dr. Jeff Dyche, Dr. John Hathcoat, and Dr. Scott Sperling, for their mentorship, constructive criticism, and commitment to my growth as a student and researcher. Additionally, I would like to thank the entire multi-disciplinary team at the University of Virginia Health System's Adult Neurology Memory and Aging Care Clinic. I am incredibly grateful to have had the opportunity to complete part of my research apprenticeship at this clinic under the mentorship of Dr. Sperling, and would not have formulated the idea for this thesis research project without applying the knowledge that I have learned from that experience. Last but not least, I would like to thank my parents, Richard Matusz and Theresa Frank, for their continuous support and unwavering faith throughout the entirety of my academic career.

Table of Contents

List of Tables

List of Figures

Abstract

Approximately 13 percent of the American population are 65 years of age or older (Vincent & Velkof, 2010). Of these 48 million older adults, roughly 5.3 million have received a clinical diagnosis of Alzheimer's disease (AD) (Alzheimer's Association, 2017). As the awareness of AD continues to heighten, so does the push for increased cognitive screening to identify signs of abnormal aging. However, important considerations pertaining to scale development or weighting procedures applied during the test development process remain unclear, as they are often not reported in testing manuals. The current study presents a statistically derived scoring algorithm for a brief screening measure of cognition, the Mattis Dementia Rating Scale 2 (DRS-2) in a sample of 113 older adults (55 Alzheimer's disease dementia, 58 Mild Cognitive Impairment. Logit weights obtained from logistic regression analysis were utilized to re-weight the subscales of the DRS-2 to reflect the order of relative importance of the five DRS-2 subscales. Sensitivity and specificity rates of the original and logit-weighted DRS-2 scores were compared to examine the impact of weighting on DRS-2 classification accuracy. Results indicated an increase in sensitivity from 78% to 90% and a decrease in specificity utilizing the newly computed logit-weighted scores. These results highlight the importance of scale construction during the instrument development process, suggesting that weighting procedures directly affect measurement utility. Additional implications for future clinical practice and research are discussed.

Keywords: dementia rating scale, neuropsychological assessment, cognitive screening, Alzheimer's disease, measurement

Screening for Dementia: An Examination of Subscale Relative Importance

I. Introduction

Dementia, an umbrella term referring to symptoms that result from different types of brain disorders, has become a familiar word in the healthcare community. With the increasing aging of the baby boomer generation, a call for increased dementia screening in primary healthcare settings and earlier, repeated administration of neuropsychological testing has been issued as the dementia continues to gain recognition as a worldwide epidemic (Boustani, Peterson, Hanson, Harris, & Lohr, 2003; Heller, Scott, Janicki, & Pre-Summit Workgroup on Caregiving and Intellectual/Developmental Disabilities, 2018; Iliffe, Manthorpe, & Eden, 2003; Imtiaz, Tolppanen, Kivipelto, & Soininen, 2014). The disease single-handedly threatens healthcare systems and economies worldwide, with an estimated $818 billion devoted to dementia care (Prince, Comas-Herrera, Knapp, Guerchet, & Karagiannidou, 2016). Heightened awareness of dementia has resulted in increased frequency in the administration of cognitive screening tests and neuropsychological evaluations to the aging population. Thus, the projected rising cost of dementia care highlights the importance of the utilization of reliable and valid measures to identify changes in cognition associated with a diagnosis of a dementia. Improved methods of detecting cognitive deficit in older adults at risk for dementia provide the patient, caregiver, and clinician with the appropriate knowledge necessary to make critical treatment decisions. In doing so, the patient's risk of experiencing decreased quality of life as the disease progresses lessens. Early detection of deficit provides a wider range of treatment options for the patient. Although there is currently no cure for dementia, patients diagnosed with AD dementia often benefit from

appropriate use of proper caregiver planning, interventions, and pharmacological treatments aimed to improve quality of life and other aspects of physical health (Liu-Seifert, Siemers, Sundell, Price, Han, Selzler… & Mohs, 2014). Accurate and time appropriate detection of cognitive deficit indicative of AD provides patients with longer access to these resources in comparison to patients diagnosed later on in the disease course. This often results in a slower progression of the course of the neurodegeneration when time-sensitive pharmacological treatments are administered appropriately, as well as an increased quality of life reported by both the patient and caregiver (Appels & Scherder, 2010; Giebel, Sutcliffe, Stolt, Karlsson, Renom-Guiteras, Sotos… & Challis, 2014). For example, research supports the efficacy of Donepezil in slowing the progression of cognitive decline associated with AD in the mild and moderate stages of the dementia, but not the severe stage (Hashimoto, Kazui, Matsumoto, Nakano, Yasuda, & Mori, 2005; Krishnan, Charles, Doraiswamy, Mintzer, Weisler, Yu, … Rogers, 2003). In contrast, research supports the efficacy of Memantine in slowing the progression of associated cognitive decline in the moderate to severe stages, but not the mild stage (Danysz, Parsons, Mobius, Stoffler, & Quack, 2000; Roundtree, Chan, Pavlik, Darby, Siddiqui, & Doody, 2009; Wilkinson, 2012). The temporal association between drug administration and treatment efficacy leaves patients diagnosed in the early stages with more treatment options in comparison to patients diagnosed in the later stages. Therefore, patients in the earlier stages are able to most efficiently take advantage of treatment options currently available, subsequently resulting in better treatment outcomes.

Although neuropsychological evaluation is considered the "gold standard" for the diagnosis of dementia, financial and time constraints may prevent many individuals access to this form of comprehensive assessment, as it is a costly and time-consuming process (Borson, Frank, Bayley, Boustani, Dean, Lin, … & Ashford, 2013; Bradford, Kunik, Schulz, Williams, & Singh, 2009). In such circumstances, access to reliable and valid screening measures of cognition becomes a vital determinant of patient health outcome. This trade off, however, comes at a cost, as screening measures are not as sensitive to the detection of change in cognitive abilities in comparison to a comprehensive neuropsychological evaluation. This is likely attributable to the brevity of such screening instruments, as there is a positive relationship between the number of questions on any given measure and the reliability of the measure (Peter, 1979). Short-form screening exams contain fewer questions than a battery of cognitive tests combined to form a comprehensive neuropsychological evaluation, resulting in decreased reliability of brief screening measures relative to that of complex exams.

Neuropsychological literature has become increasingly flooded with scholarly efforts to improve early detection methods of AD dementia, the sixth leading cause of death in the United States (Kochanek, Murphy, Xu, & Tejada-Vera, 2016; Nestor, Scheltens, & Hodges, 2004). These efforts are paralleled by additional efforts to improve the overall quality of neuropsychological assessment across various clinical populations (Chaytor, Schmitter-Edgecombe, & Burr, 2006; Mitrushina, 2009; Ramsay & Reynolds, 2000; Vanderploeg, 2014). Despite vast efforts to improve the validity of measures intended to assess cognitive ability, neuropsychological measurement continues to receive criticism (Meyer, Finn, Eyde, Kay, Moreland, Dies, Eisman, & Reed, 2001;

Retzlaff & Gibertini, 2014; Smith, Breitbart, & Platt, 1995). More recently, authors of a recent study posited that the proportion of subtest scoring of a brief screening measure of cognition, the Montreal Cognitive Assessment (MoCA), appears unreasonable (Fengler, Kessler, Timmermann, Zapf, Elben… & Kalbe, 2016). Commonly used as a screening tool for dementia, Fengler et al. (2016) suggest the current scoring method of the MoCA does not align with empirical support for the rank ordering of the cognitive constructs of greatest to least importance when used to screen for abnormal cognitive impairment in patients with Parkinson's disease dementia. Fengler and colleagues (2016) re-weighted these subscale scores according to the Areas under the Curve (AUC) in a Receiver Operating Characteristic (ROC), and reported a 29.5% increase in sensitivity and a 4.7% decrease in specificity. Ultimately, the authors of this study highlighted the effect of differential subtest weighting on test utility, bringing attention to the potential negative effects that may be seen if proper consideration is not given to these aspects of measurement development (Fengler et al., 2016). This remains to be one of the only studies in the field of neuropsychology examining the effects of subscale weighting. This calls attention to the dearth of literature investigating whether or not changes can be made to the structure of neuropsychological tests to improve instrument utility in the clinical setting. Furthermore, little has been published in the field of neuropsychology regarding the examination of the quality of test development procedures. With little to no attention devoted to an investigation of the current standards of these procedures, it remains unclear as to how these assessments are being created, and why they are being created in the ways that they are. For example, due to the under-reporting of these processes in formally published testing manuals, it may remain unclear as to why a test developer

devoted ten questions to the assessment of one cognitive construct, but only five questions to another. Why is one subscale worth more points than another, when the two subscales contain the same number of questions? Because these decisions are not explicitly reported in the measurement development section of neuropsychological testing manuals, one is unable to assume whether appropriate consideration was given to these foundational components of the measurement development process. Ultimately, it appears plausible that some of the posited issues related to neuropsychological assessment may be deeply rooted within the practices with which these measures are initially constructed. Thus, reactive efforts aimed at improving measures after they have already been created and disseminated to practitioners for use may ultimately prove sub-par, such that revision efforts at this stage may ultimately be considered, "too little, too late." As such, attempts to increase the reliability of a measure through minor revisions of measurement content may fail to address an underlying problem: the construction of the measure itself. Neuropsychological instrument development must resemble a practice firmly grounded in measurement theory in order for these tests to reach full capacity in terms of reliable and valid clinical utility.

It is important to note that this may be the case only if test developers are not providing appropriate consideration to development procedures during measurement construction. This remains unclear, however, as many neuropsychological instruments fail to report important considerations pertaining to measurement construction during the test development process in published testing manuals. For example, the testing manuals for two commonly used brief cognitive screening instruments, the Mattis Dementia Rating Scale 2 (DRS-2) and the Mini Mental State Examination (MMSE) both lack

detailed information regarding the test development process. Specifically, it is unclear how and why items were generated and selected for inclusion. No information is provided detailing the creation of the scoring methods utilized for both instruments, and no reasoning for the use of the current scoring method is included in either of these testing manuals. It is important to note that the test authors may, in fact, have provided appropriated consideration to these factors during the test development process. However, that possibility remains unclear, due to the exclusion of a description of this consideration, should it have occurred during the initial construction of the instruments. Insufficient attention to and reporting of measurement scale development (i.e., the determination of subscale weights) may leave these measures and the interpretation of their scores vulnerable to error, especially when proposed cut scores are utilized to inform clinical decisions. Depending on the degree to which subscale score weights differ, poor performance on a lower-weighted subscale will differentially affect the total score of the measure more than performance on a higher-weighted subscale. Increased variability among subscale weights may render an individual more susceptible to receiving a total score equal to or below the recommended cut score associated with the measure, if his or her performance shortcomings are predominantly evidenced on subscales maintaining a higher weight, relative to the other subscales within the same measure. In contrast, a wider range of subscale weights may render an individual more susceptible to receiving a total raw score above the recommended cut score, if principal shortcomings are evidenced on lower-weighted subscales, relative to other subscales. For example, if a measure worth a total of 20 points contains give subscales and maintains a recommended cut score of 17/20 total points, the use of differentially weighted subscale

scores will result in the subscale with the greatest weight being most influential on the outcome of the total raw score, relative to the cut score comparison point. If four of the subscales are scored out of a three possible points, whereas the fifth subscale is scored out of eight possible points, getting 50% of the questions correct on the eight-point subscale would result in the test-taker falling below the recommended cut score. In comparison, getting 50% of the questions correct on any one of the three-point subscales subscales would result in a score above the recommended cut score for the test-taker. Ultimately, the total raw score may not appropriately reflect an underlying 50% deficit in one area of cognition if a-theoretical differential weighting methods are applied to the subtests of an instrument.

Test developers should give appropriate consideration to the measurement development process, and decisions involved in this process should be based on empirical or statistical grounds (Baldwin, 2015). Given the additive nature of composite scores, researchers have historically devoted significant attention to the determination and analysis of methods aimed to optimize the weighting of subscales (Govindarajulu, 1988; Perloff & Persons, 1988; Thompson, 1940; Wang & Stanley, 1970). Govindarajulu (1988) proposed that the assignment of differential weights to composite components during the measurement construction process might not accurately reflect true weights. He further argued that this will most likely not remain true when scores are highly related to one another if appropriate consideration was not devoted to this source of potential influence during measurement development (Govindarajulu, 1988). Research also suggests that the use of incorrectly weighted indices introduces bias into the measurement of a given criterion (Perloff & Persons, 1988). Perloff and Persons (1988) further

supported this claim, reporting a notable 33% increase in the estimated explanatory power of a measure when comparing results derived from unjustified equally-weighted indices to results derived from differentially-weighted indices that were empirically and statistically justified. Results from the study conducted by Perloff and Persons (1988) exemplified how the use of incorrectly weighted indices can leave a test vulnerable to error by introducing bias into criterion measurement, as evidenced by differences in the magnitude of the effect associated with the predictive utility between weighting conditions.

Baldwin (2015) raised concern regarding the use of a priori or subjective weights, as explanation is rarely provided regarding the grounds on which these decisions are made. Burt (1950) explained that this weighting procedure involves the alignment of the differential contribution of subjective weights and the experts' beliefs regarding the relative importance of each of the subscales. Baldwin (2015) argued that experts fail to address important empirical and statistical considerations regarding the variances of the subscales, covariance among the subscales, and the reliability of each of the subscales when assigning weights to subscales. Fluctuating beliefs regarding subscale construct validity, as well perceived differences regarding subscale length and item difficulty is often reflected in corresponding beliefs regarding the relative importance of subscales. Therefore, the subscale weights more appropriately reflect the experts' beliefs about those aforementioned factors, as opposed to the importance of the sub-constructs with respect to the actual measurement of the intended outcome (Baldwin, 2015). In sum, Baldwin (2015) pushed for an improvement in the reporting of measurement development procedures by test authors.

The conversation of measurement principles in the field of neuropsychology, however, remains somewhat limited. For example, a predominant topic of conversation in the field pertains to the sensitivity-specificity trade-off, such that neuropsychologists have direct substantive efforts towards improving clinical test utility. Sensitivity is best defined as, "the proportion of people with the condition of interest correctly identified by a test as having the condition" (Lezak, Howieson, Bigler, & Tranel, 2012). Specificity, on the other hand, is described as, "the proportion of people without the condition of interest correctly identified by a test as not having the condition" (Lezak et al., 2012). There is a trade-off between sensitivity and specificity, such that no test can have a perfect degree of both. Ultimately, a test that is more sensitive lends itself to an increased number of "false alarms," or false positive cases. In contrast, a test with greater specificity than sensitivity lends itself to fewer false positive cases, but also a decreased number of true positives identified. When considering the sensitivity-specificity trade-off, test developers and researchers alike must determine which is more important given the purposes of a test.

Sensitivity is often highly valued for the purposes of cognitive screening tests, as the benefits of allowing for more accurate detection of true positive cases, as evidenced by increased sensitivity, outweigh the risks associated with a false positive score. Subsequent decreases in specificity would then result in an increased number of Type II errors, potentially resulting in increased financial expenses, as patients who fall below cut scores on cognitive screening tests are then referred for a comprehensive neuropsychological evaluation. However, even if the results of the comprehensive test indicate normal levels of cognitive functioning, the costly testing session will then serve

as a normative baseline for future comparative purposes, should concerns of deficit arise at a later point. Therefore, many researchers dedicate time and effort towards researching ways to increase the sensitivity of cognitive screening measures for dementia, as decreasing Type II error by increasing measurement sensitivity is more highly valued than the decreasing of Type I error that would be associated with increased measurement specificity.

The current study attempted to address these issues by evaluating the relative utility of the subscales of a brief cognitive screening instrument for dementia. The primary goal of this study was to determine if the rank ordering of statistically generated subscale weights align with the rank ordering of the initial subscale weights applied during the original measurement construction process. A secondary goal of the current study was to examine whether utilization of a statistically supported subscale-weighting algorithm could improve the sensitivity of a brief cognitive screening measure used to identify abnormal levels of cognitive impairment consistent with a diagnosis of dementia in older adults, the Mattis Dementia Rating Scale 2, above and beyond the sensitivity rate obtained utilizing the original scoring method.

II. Review of the Literature

2.1 Cognitive Aging

2.1.1 Age-related cognitive decline (ARCD). Understanding cognitive functioning during the human lifespan, especially in early development or old age, can provide unique insight into a person's physical and psychological well-being. The development of neuropsychological tests designed to measure overall cognition or individual aspects of cognition (e.g., memory, visuospatial abilities, learning, processing

speed, attention) has afforded clinicians the ability to quantitatively measure patient cognitive performance. In the context of older adults diagnosed with neurodegenerative dementia, initial fluctuations in some previously stable area of cognitive functioning, as evidenced by below average neuropsychological testing scores, often serve as the principal indicator of the proliferation of neuronal cell death associated with the disease (Palmer, Backman, Winblad, & Fratiglioni, 2003). However, this may not always be the case, as clinical neuropsychologists are faced with the challenge of distinguishing changes in cognitive abilities that result from typical age-related processes from changes resulting from an atypical pathological process, such as a neurodegenerative dementia.

Increasing age correlates strongly with an increasing susceptibility to changes in cognition. Older adults will inevitably experience a normal degree of slowing in the speed of mental processing, often resulting in diminished performance across various domains of cognition, including memory, attention, language, and executive functioning abilities (Smith & Bondi, 2008). Several theories of aging have been proposed in attempts to explain the reasoning for the slowing of mental processes associated with normal aging. Generally, most modern aging theories fall within one of two overarching theoretical categories: programmed or damaged theories.

Programmed theories of aging propose that aging, similar to earlier stages of development, is an unavoidable process that is pre-determined by a biological clock (Goldsmith, 2012; Jin, 2010). Essentially, these theories propose that age-related processes are simply an inevitable part of lifespan development. Programmed theorists have also attributed an increased susceptibility to illness in older adults to an immune system that is programmed to decline with increasing age (Kowald & Kirkwood, 2016).

Ultimately, these researchers have attributed this this immune system vulnerability to the declines in health and cognition in older adulthood. Another programmed theory of aging attributes cognitive decline in later life to programmatically timed genes that "turn on" in older adulthood, specifically.

Damaged theories of aging have gained vast popularity since programmed theories have generally been discredited (Blagosklonny, 2013; Maynard, Fang, Scheibye-Knudsen, Croteau, & Bohr, 2015). Generally, these theories attribute age-related declines to a biological catalyst or dysfunction, which ultimately results in damage to the brain, rendering it unable to continue to operate optimally and efficiently (Kirkwood & Melov, 2011).

Despite theoretical discrepancies, there is general agreement among researchers that age-related cognitive changes, to a degree, are associated with normal, non-pathological processes and are believed to be benign in nature. The presentation of cognitive decline due to pathological neural processes is often, in the early stages, very similar to age-related cognitive decline. The etiological underpinnings of this atypical decline, however, remain unclear.

2.1.2 Mild cognitive impairment (MCI). Petersen and colleagues adopted the term mild cognitive impairment (MCI) to describe declines in cognitive functioning that extend beyond the changes associated with age-related cognitive decline (ARCD), or cognitive decline associated with normal aging, but do not quite warrant a diagnosis of dementia (Petersen, Smith, Ivnik, Tangalos, Shaid, Thibodeau... & Kurland, 1995). The original MCI diagnostic criteria included the presence of a subjective memory complaint, below average performance on memory tasks, preservation of general cognitive

functioning, maintenance of normal activities of daily living (ADLs), and the preservation of global cognition (Petersen, Smith, Waring, Ivnik, Tangalos, & Kokmen, 1999). Petersen, Caracciolo, Brayne, Gauthier, Jelic, and Fratiglioni (2014) proposed more recent diagnostic guidelines for MCI, which provide additional specification of the criteria in order to allow for improved clarification of the distinction between a diagnosis of MCI and dementia. Similar to the original diagnostic criteria for MCI, current criteria includes the presence of a subjective complaint of cognitive impairment and maintenance of independence of ADLs (Petersen et al., 2014). This newly adopted criteria, however, includes the objective determination of impairment in one or more domains of cognition (Petersen et al., 2014). Additionally, the current criteria assert that the changes in cognition are not significant enough to hinder functioning in social or occupational contexts, precluding a diagnosis of dementia (Petersen et al., 2014).

2.1.3 Dementia. Dementia is not a specific disease, but rather an umbrella term that refers to symptoms resulting from different types of brain disorders (World Health Organization, 1992). The term "dementia" broadly categorizes many different diseases, all of which are characterized by a marked decline in baseline cognitive functioning that impedes a person's ability to complete ADLs (Schoenberg & Duff, 2011). The primary distinction between MCI and dementia is the independent completion of ADLs. Both a person with MCI and a person with dementia will exhibit impairments in cognitive abilities on tests of cognition when compared to age and education-based normative data. A person with dementia, however, will require assistance with the completion of one or more ADLs, whereas a person with MCI will maintain the ability to complete ADLs independently. Longitudinal studies have reported annual conversion rates from MCI to

dementia, ranging from approximately 2-31% (Mitchell & Shiri-Feshki, 2009; Bruscoli & Lovestone, 2004; Fischer, Jungwirth, Zehetmayer, Weissgram, Hoenigschnabl, Gelpi … & Tragel, 2007). The conversion rate increases significantly when participants are monitored for a period of time that extends beyond one year. When monitored over a two-to-three year period, conversion rates significantly increased, ranging from approximately 26-84% (Aerts, Heffernan, Kochan, Crawford, Draper, Tollor, Sachdev, & Brodatv, 2017; Fischer et al., 2007; Petersen et al., 1995; Petersen, Stevens, Ganguli, Tangalos, Cummings, & DeKosky, 2001; Tabuas-Pereira, M., Baldeiras, I., Duro, D., Santiago, B., Ribeiro, M.H., Leitao, M.J., Oliveira, C., & Santana, I., 2016). Despite this variability across studies, there is a general consensus among researchers that MCI is considered a risk factor for dementia (Risacher, Saykin, West, Shen, Firpi, McDonald, & the ADNI, 2009). Regardless of etiology or the course of progression of the disease, the fundamental feature of all dementias is diminished cognitive ability (American Psychiatric Association, 2000). This primary deficit has remained the focal point for dementia diagnoses across all iterations of diagnostic guidelines that have cycled through the research and clinical fields.

Philippe Pinel, known as the father of modern psychiatry, is commonly credited with being one of the first to introduce the term d**é**mence, meaning dementia, to the scientific community through his detailed descriptions of the disease in 1797 (Boller & Forbes, 1998). Nearly two centuries later, the term appeared for the first time in the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM), where it was defined as, “a loss of intellectual abilities of sufficient severity to interfere with social or occupational functioning” (American Psychiatric Association, 1980). In a more

recent shift, the DSM-5 introduced the term *major neurocognitive disorder* (major-NCD), under which the term dementia is subsumed (American Psychiatric Association, 2013). Major-NCD is characterized by a subjectively reported and objectively evidenced decline in at least one cognitive domain, resulting in interference in a person's ability to function independently (American Psychiatric Association, 2013). To warrant a diagnosis of major-NCD, the cognitive decline must have resulted from something other than a delirium or another mental disorder (American Psychiatric Association, 2013). The DSM-5 diagnostic criteria also include behavioral, severity, and etiological specifiers for major-NCD. The clinician will specify whether a patient's cognitive decline is or is not accompanied by behavioral disturbance. Secondly, the clinician will specify the current severity of the patient's cognitive decline as mild, moderate, or severe. Major-NCD of mild severity is characterized by difficulties with instrumental ADLs (American Psychiatric Association, 2013). These activities refer to more complex activities that require higher levels of processing, such as financial management. Major-NCD of moderate severity is characterized by difficulties with basic ADLs (American Psychiatric Association, 2013). These activities refer to the more simplistic everyday tasks necessary for fundamental functioning, such as going to the bathroom or bathing. Persons experiencing difficulties only completing instrumental ADLs require much less assistance from others in comparison to persons experiencing difficulties completing basic ADLs. Specifically, the loss of independence in the completion of basic ADLs therefore implies the loss of independence in the completion of instrumental ADLs. For example, a person requiring assistance with routine hygienic care due to cognitive impairment would correspondingly require assistance with financial management.

However, both persons with mild and moderately severe major-NCD maintain some degree of independence. In contrast, persons with severe-stage major-NCD are fully reliant on others for all ADLs. Lastly, clinicians will specify whether the dementia is due to any of the following causes: Alzheimer's disease, Frontotemporal lobar degeneration, Lewy body disease, vascular disease, traumatic brain injury, substance/medication use, HIV infection, Prion disease, Parkinson's disease, Huntington's disease, another medical condition, multiple etiologies, or an unspecified etiology (American Psychiatric Association, 2013).

The Clinical Manual for the 10th edition of the International Classification of Disorders (ICD-10) also includes a definition and description of dementia diagnostic criteria, referred to as "unspecified dementia" in diagnostic code F03. The ICD-10 describes dementia as, "a condition in which a person loses the ability to think, remember, learn, make decisions, and solve problems" and as, "an acquired organic mental disorder with the loss of intellectual abilities of sufficient severity to interfere with social or occupational functioning" (World Health Organization, 1992). For health insurance purposes, clinicians' reference specific codes listed for clinical conditions described in diagnostic manuals when providing a formally written account of a patient diagnosis. Therefore, the diagnostic criteria within these manuals are predominantly utilized within the health insurance system, as these codes allow for insurances to efficiently deem patients eligible or ineligible for certain coverage and reimbursement. However, physicians typically do not reference these manuals directly to make formal clinical diagnoses. Instead, clinical neuropsychologists typically refer to empirically supported literature and published recommendations for updated guidelines created by

carefully constructed task forces comprised of some of the most well versed researchers and clinicians on the related topic in the scientific community. The codes presented in the diagnostic manuals are primarily referenced for billing and insurance purposes.

In 1984, the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) work group described dementia syndrome as "the decline of memory and other cognitive functions in comparison with the patient's previous level of function as determined by a history of decline in performance and by abnormalities noted from clinical examination and neuropsychological tests" (McKhann, Drachman, Folstein, Katzman, Price, & Stadlan, 1984). These criteria specify that behavior serves as the basis for a diagnosis of dementia and that a diagnosis cannot be made in the presence of another medical condition that impacts cognition, as it would invalidate cognitive test results vital to the diagnostic procedure due to the inability to determine the true origin or cause of cognitive deficit.

More recently, workgroups from the National Institute on Aging-Alzheimer's Association (NIA-AA) proposed a set of core clinical criteria for "all-cause dementia" that currently reflects the most widely accepted and referenced guidelines for dementia diagnosis (McKhann, Knopman, Chertkow, Hyman, Jack, … Phelps, 2011). These criteria indicate that a diagnosis of dementia is warranted in the presence of "neuropsychiatric symptoms that interfere with the ability to function at work or at usual activities, represent a decline from previous levels of functioning and performing, and are not explained by delirium or major psychiatric disorder" (McKhann et al., 2011). McKhann et al. (2011) also specify that the detection and diagnosis of cognitive deficit

must occur via a clinical interview involving the patient and a competent informant in order to thoroughly review patient history information, as well an objective assessment of cognitive ability obtained from the administration of brief cognitive testing or a neuropsychological evaluation. Furthermore, the neuropsychiatric sequela must involve two or more of the following: "impaired ability to acquire and remember new information; impaired reasoning and handling of complex tasks, poor judgment; impaired visuospatial abilities; impaired language functions (speaking, reading, writing); or changes in personality, behavior, or comportment" (McKhann et al., 2011).

2.1.4 Alzheimer's disease (AD). Alzheimer's disease, the most common form of dementia, was initially identified in 1906 when Dr. Alois Alzheimer identified distinct post-mortem brain abnormalities of a woman who had died due to what was then considered an unexplainably rare mental illness. This woman, known as Auguste D, was described to have experienced atypical behavioral disturbances and significant declines in cognitive functioning, specifically with respect to her memory and language abilities. It was through this post-mortem study that Dr. Alzheimer would receive credit for the identification of a disease that researchers have since coined, *Alzheimer's disease (AD)*. The presence of two distinct brain abnormalities, neurofibrillary tangles and beta amyloid plaques, identified by Dr. Alzheimer remain the hallmark of the neurodegenerative disease.

Several diagnostic criteria for AD have circulated, and continue to circulate, throughout the scientific community. McKhann et al. (1984) also proposed diagnostic criteria for AD in addition to the previously described proposed dementia diagnostic criteria. An updated version of these criteria outlines the current, most widely referenced

guidelines for AD diagnosis, proposing three distinct classifications of AD diagnosis: *probable AD dementia*, *possible AD dementia*, and *probable AD dementia with evidence of the AD pathophysiological process* (Mckhann et al., 2011). Distinctions are made between the three terms based on the information obtained by the clinician during the diagnostic workup. All three terms classifying an AD diagnosis rely on the preceding determination that the patient meets criteria for all-cause dementia.

McKhann et al. (2011) indicated that a diagnosis of *probable AD dementia* is warranted when there is an insidious onset of cognitive deficit, a history of progressive cognitive decline clearly determined by means of subjective and informant reports or observations, and a either an amnestic or non-amnestic presentation of the initial and most prominent deficits as determined by history or objective assessment. Most commonly, AD presents as amnestic, in which initial and most prominent cognitive deficits are in learning and memory recall (McKhann et al, 2011). Cases in which deficit is initially and predominantly noted in language functioning (e.g., word-finding difficulties), visuospatial abilities (e.g., difficulties with spatial cognition, facial recognition), or executive functioning (e.g., reasoning, problem-solving) define the less common, non-amnestic presentation of AD (McKhann et al., 2011). AD is a differential diagnosis, in which all other possible causes of cognitive decline must first be ruled out in order for a diagnosis of AD to be made.

In contrast to a diagnosis of *probable AD dementia*, a diagnosis of *possible AD dementia* is appropriate when the course of patient deficits presents atypically if the patient either does not evidence an insidious onset of cognitive decline, or the clinician lacks a detailed patient history and is therefore unable to identify a clear pattern of

progressive decline (McKhann et al., 2011). Possible AD dementia may also be diagnosed in cases in which the patient has a history of cerebrovascular disease, evidence of Dementia with Lewy Bodies, or evidence that the decline in cognitive functioning may have resulted from an alternative factor (McKhann et al., 2011). McKhann et al. (2011) classify these cases as representing an "etiologically mixed presentation" of deficit, in which the clinician is unable to parse apart whether or not the sole cause of cognitive deficit can be solely attributed to AD pathology.

McKhann et al. (2011) presented a third term, *probable AD dementia with evidence of the AD pathophysiological process*, for the classification of AD diagnosis. Evidence for this process may be obtained through the measurement of identified biomarkers specific to AD, which are pathogenic biological components contained in in the blood, cerebrospinal fluid (CSF), or other tissue (Loring, 2015). Biomarkers specific to AD include the presence of amyloid-beta protein and downstream neuronal degeneration (McKhann et al., 2011). The amino acid peptide, amyloid-beta is produced by the amyloid precursor protein (APP) and is found within the amyloid plaques that characterize AD (Loring, 2015). Abnormal depositions of amyloid-beta can be identified via CSF analysis or amyloid positron emission tomography (PET) (Loring, 2015; McKhann et al., 2011). This neuroimaging technique is used to identify the presence of amyloid-beta neuritic plaques in the brain (Loring, 2015). AD is one of more than 20 clinicopathological entities, collectively referred to as tauopathies (Williams, 2006). Therefore, abnormal elevations of phosphorylated tau (p-tau), total tau, and CSF tau also serve as biomarkers of AD (McKhann et al., 2011). McKhann et al. (2011) presented this third term for to more specifically classify a diagnosis of possible and probable AD, as

the presence of AD biomarkers increases diagnostic certainty that the cognitive decline is truly attributable to the pathophysiological process seen in AD. These pathogenic distinctions are believed to be responsible for the neurodegeneration and associated decline in cognitive functioning seen in AD.

2.2 Neuropsychological Evaluations and the Diagnosis of Dementia

The appearance of widespread cognitive deficits of increased severity, beyond those expected with normal aging, result from manifestations of pathological abnormalities that preclude optimal neurological functioning. Diminished cognitive capacities of older adults result in notable changes in behavioral and functional abilities. Changes associated with age-related cognitive decline (ARCD) are generally subtle and non-debilitating in nature. In contrast, changes associated with a neurodegenerative dementia, such as AD, are more severe and long lasting and result in decline that excels far beyond the expected changes in cognition associated with normal ARCD. Neuropsychologists are faced with the challenge of differentiating changes in cognition due to ARCD from changes in cognition that result from atypical neurodegenerative processes.

A person experiencing symptoms of cognitive decline is most typically referred for a neuropsychological evaluation after family members or caregivers report changes in the person's cognitive abilities. Neuropsychologists administer these comprehensive evaluations in order to make inferences about brain-behavior relationships, make diagnoses, and describe relative strengths and weaknesses to aid treatment and rehabilitation. The neuropsychological evaluation is typically comprised of a clinical interview, a report of the patient's history, and cognitive assessment (Lezak et al., 2012).

2.2.1 Clinical interview. A comprehensive review of qualitative patient data is a key component in the neuropsychological evaluation. The clinical interview provides the neuropsychologist with a more comprehensive understanding of the patient's cognitive and functional abilities. The information from the clinical interview places the data obtained from cognitive testing into context, allowing the neuropsychologist to best draw conclusions regarding the brain-behavior relationships of a given patient. The clinical interview involves the collection of information pertaining to the patient's social, occupational, educational, medical, and psychiatric history, presenting problems, ADLs, and the current status of each component from both the patient and a secondary informant. More often than not, individuals with dementia lack insight with respect to their presenting deficits. In a study conducted by Shany-Ur, Lin, Rosen, Sollberger, Miller and Rankin (2014), degree of lack of awareness was highly positively correlated to neural atrophy in regions of the brain responsible for attention, self-related processing, and of self-knowledge. Wilson, Boyle, Yu, Barnes, Sytsma, Buchman, Bennett, and Schneider (2015) reported that lack of awareness of memory impairment, specifically, appeared two to three years prior to the onset of the dementia, on average. For this reason, someone closely related to or closely involved with the care of the individual with dementia, serves as a secondary informant during the clinical interview process. A patient and the informant may be interviewed together or separately. For example, a patient exhibiting signs of increased frustration or aggression may be interviewed separately from the informant in order to decrease levels of patient or informant distress. A primary goal of this component of the neuropsychological evaluation is to obtain information about the presenting symptoms. Specifically, questions are often asked to

help identify when and how the symptoms began, as well as the ways in which the described symptoms may have changed over time.

2.2.2 Record review and neuroimaging. The clinician will review relevant patient records of medical, psychiatric, and social history in order to determine if the presenting symptoms could solely due to, or an exacerbation of, preexisting causes other than neurodegenerative processes that typify dementia. This review is necessary as dementia is a differential diagnosis requiring all other possible differentials to first be ruled out in order to a clinical diagnosis of dementia to be concluded. When available, the neuropsychologist will obtain and review the patient's medical records. Most often, these are obtained from the source through which the referral for the neuropsychological evaluation was made. These records most typically provide information regarding current medications, medical workups, and medical history.

In order to arrive at a clinical diagnosis, the clinician must first rule out all other possible causes for the dementia. Specifically, it is important to first distinguish if the underlying cause of the dementia-related symptoms is the result of an alternate and potentially reversible etiology. For example, depression, toxic encephalopathy, and even some rare diseases, can result in cognitive disarray that mimics dementia symptomology (Schoenberg & Scott, 2011; Singer, 2011). Through a comprehensive review of the patient's medical records, the clinician is able to extract relevant information to identify risk factors and evidence for alternative causes of the presenting symptoms. Results from neuroimaging methods are also often gathered to provide additional information aiding differential diagnosis during the initial diagnostic workup.

Neuroimaging techniques provide valuable information for the differential diagnosis of dementia. Magnetic resonance imaging (MRI) scans are essential for determining the current structural nature of the brain. A structural imaging method, MRI allows for the identification of volumetric changes of the brain that are consistent with the progression of neurodegenerative diseases (Tartaglia, Rosen, & Miller, 2011). Changes in degree of ventricular volume, global brain volume, and bilateral basal ganglion white matter have been linked to neurodegenerative progression represented by the conversion from MCI to AD (Li, Tan, Wang, Tan, Tan, Xu, Zhao, Wang… and Yu., 2016). It has been suggested that the amount of atrophy identified within Medial Temporal Lobes serves as the most reliable MRI indicator of conversion from MCI to dementia (Risacher et al., 2009). As neurodegenerative processes associated with AD continue, atrophy expands beyond the Medial Temporal Lobes, ultimately resulting in global atrophy across all brain regions (Risacher et al., 2009). The degree of global atrophy is most typically highly related to the severity of the stage of dementia, such that higher degrees of global atrophy are most commonly seen in the later stages of dementia. Neuroimaging alone, however, is not sufficient for a diagnosis of dementia.

2.2.3 Cognitive assessment. Neuropsychologists utilize information obtained from the clinical interview and patient history in tandem to aid the interpretation of cognitive assessment, which serves as an objective measurement of cognitive ability across domains of cognition. A typical neuropsychological test battery for individuals with dementia involves the assessment of five cognitive domains: Attention and executive functioning, learning and memory, language and speech, visuoperception and visuoconstruction, and emotion, mood, and personality (Schoenberg & Scott, 2011). The

neuropsychologist utilizes a combination of both nomothetic and ideographic approaches when interpreting test results, comparing the quantitative assessment data to age and education based norms, and previously obtained quantitative data indicative of individual baseline functional ability, if previously obtained. These comparisons allow for the identification of deviations from "normal" levels of cognitive functioning. Repeated administration of neuropsychological testing allows for the tracking of these patterns of cognitive ability over time, facilitating the identification and analysis of deviations from nomothetic and ideographic norms. Researchers have examined these trends of cognitive deficits across clinical dementia populations, from which they have deduced "typical" cognitive profiles of different types of dementias. Ultimately, clinicians rely on the identification of patterns of cognitive deficit when differentiating among different types of dementia (John, Gurnani, Bussell, Saurman, Griffen, & Gavett, 2016; Ritter, Leger, Miller, & Banks, 2017). Only after exhausting all available resources to obtain information required to rule out all other possible causes of the patient's decline in cognition, will a clinician provide a patient with a clinical diagnosis of dementia.

Neuropsychologists will either administer either a fixed or flexible battery of cognitive tests. A *fixed battery* involves the administration of a preselected grouping of cognitive tests. For example, a neuropsychologist utilizing this approach would administer the same exact battery to all patients presenting with difficulties related to cognitive functioning regardless of individual differences, such as the severity of impairment, rate of decline, or the presentation of the deficits. Essentially, these batteries adopt a "one size fits all" approach. In contrast, a flexible battery adopts an individualized approach, which has gained vast popularity in current, Western

neuropsychological practice. As noted by Kubu, Ready, Festa, Roper, and Pliskin (2016), most North American neuropsychologists report a strong preference for the utilization of flexible batteries for the purposes of optimizing patient-centered accommodations, reducing the administration of redundant assessments, and emphasizing timeliness for both the patient and the clinician.

The *flexible battery* approach to neuropsychological assessment rejects the "one size fits all" approach, and is instead comprised of a group of cognitive tests that have been hand-selected by the neuropsychologist. The neuropsychologist determines which tests he or she will include within a flexible battery strictly based off of the needs of the individual. For example, Patient A reported increasing word-finding difficulties over the past year during the clinical interview. Additionally, the patient has a family history of AD. Using a flexible battery approach, the neuropsychologist would likely create a battery that includes more measures assessing Patient A's memory, as well as expressive and receptive language abilities. On the other hand, Patient B reported increased difficulty when writing thank you cards during the holidays. According to the review of patient history, this patient's father was diagnosed with Parkinson's disease. Based off of this information, the neuropsychologist would likely include more tests assessing Patient B's visuoconstructional, spatial, and motor abilities within the flexible neuropsychological battery.

2.3 Mattis Dementia Rating Scale 2 (DRS-2)

The Mattis Dementia Rating Scale 2 (DRS-2) is a test of general cognitive ability that can be administered either within a comprehensive neuropsychological battery or on its own as a brief screening measure of cognition. Stephen Mattis (1976) developed the

original version of the Dementia Rating Scale (DRS) to assess and track cognitive abilities in cognitively impaired populations, primarily those diagnosed with degenerative diseases. A plethora of research published after the publication of the original version of the test highlighted the effects of age and education levels on DRS test scores, warranting the publication of an updated version of the test (Mattis, 1988). Specifically, prior research has consistently demonstrated that older adults with lower levels of education perform more poorly on neuropsychological tests, which results in a misclassification of these individuals as having dementia when, in fact, the lower test scores are attributable to lower education status as opposed to a neurodegenerative disease (Marcopulos, McLain, & Giuliano, 1997; Mattis, 1988; Stern, Andrews, Pitman, Sano, Tatemichi, Lantigua, & Mayeux, 1992). In addition, research regarding the influence of age on neuropsychological test scores has reliably revealed a negative relationship between age and neuropsychological test scores, which often leads to a misinterpretation that the lower test scores result from an underlying pathology, when that may not necessarily be the case (Marcopulos, et al., 1997; Nadler, Mittenberg, DePiano, & Schneider, 1994; Miller, Myers, Prinzi, & Mittenberg, 2009; Scheiber, Chen, Kaufman, & Weiss, 2017; Wisdom, Mignogna, & Collins, 2012).The second version of the DRS was introduced in 2001 in order to provide a wider range of age and education based norms in comparison to the norms provided in the original version, which were obtained from a relatively small sample of younger adults (Mattis, 1976, 1988).

The DRS-2 identically resembles the DRS, but also includes new standardized normative data, additional interpretive guidelines linking test scores with the test norms, and a review of research published since the publication of the original version of the

DRS highlighting the reliability and validity of the measure. The norms included in the second version were derived from the Mayo's Older Americans Normative Studies (MOANS), a project designed to create age-based norms for adults ages 55 and older and to validate the clinical utility of a collection of cognitive tests (Ivnik, Malec, Tangalos, Petersen, Kokmen, & Kurland, 1990; Ivnik, Smith, Tangalos, Petersen, Kokmen, & Kurland, 1991; Ivnik, Malec, Smith, Tangalos, Petersen, Kokmen, & Kurland, 1992; Lucas, Ivnik, Smith, Bohac, Tangalos, Kokmen, Graff-Radford, & Petersen, 1998; Lucas, Ivnik, Willis, Ferman, Smith, Parfitt, Petersen, & Graff-Radford, 2005). The MOANS data were co-normed across measures to allow for the comparison of individual performance in various cognitive domains across different levels of age (Lucas et al., 1998; Lucas et al., 2005). The MOANS battery from which this expanded normative data were generated included the original DRS. The DRS-2 testing manual also includes the regression equation created by Lucas et al. (1998) adjusting for differences in education level. The expanded norms introduced in the second version of the DRS were provided to more adequately account for the variability in patient performance across age and education and therefore improve clinical interpretations of obtained scores on the measure. Identical to the original version of the DRS, the DRS-2 also consists of 36 items that are intended to assess general cognitive ability across five subscales: Attention (ATT), Initiation/Perseveration (IP), Conceptualization (CONC), Memory (MEM), and Construction (CONST). The number of possible points per subscales ranges from 6 to 39 points.

The items on the DRS-2 are not scored independently, as items within each subscale are arranged hierarchically in terms of item difficulty, with more difficult items

appearing first within each subsection of the test. If a person answers the initial and therefore most difficult questions correctly within a subsection of the test, the person will skip and automatically receive points the remainder of the questions in that section. Lower subscale scores represent lower levels of domain-specific cognitive abilities and higher levels of domain-specific cognitive deficit. Similarly, lower total DRS-2 scores represent lower levels of cognitive abilities, or a greater level of cognitive deficit. According to a study conducted by Montgomery and Costa (1983), DRS-2 total scores fell below 123/144 in 62% of patients with dementia, whereas only 12% of non-demented patients received a DRS-2 total score of 123/144. Vangel and Lichtenberg (1995) reported an 87% correct classification rate with 85% sensitivity and 90% specificity utilizing a 125 cut score in a younger sample of older adults (ages 62-79), and reported similar rates utilizing a cut score of 123 in a sample of older adults ages 80-95. The most widely referenced cut score for the DRS-2 currently remains a total score of 123/144.

Research suggests that poor performance on the Memory and Initiation/Perseveration subscales correlates highly with increased levels of cognitive impairment associated with AD (Monsch, Bondi, & Salmon, 1995). Additional research has identified a relationship between performance across different cognitive domains and ADLs. Specifically, poor performance on measures of processing speed (Owsley, Sloane, McGwin, & Ball, 2002) and executive function (Marshall, Rentz, Frey, Locascio, Johnson, & Sperling, 2012; Martyr & Clare, 2012) have been linked to poor performance on tests of everyday competence. Although processing speed is not directly assessed via the DRS-2, the initiation/perseveration subscale assesses various aspects of executive functioning, as the tasks within this subscale evaluate the ability to appropriately initiate,

alternate, and terminate behavior. Furthermore, it has been posited that poor performance on tests of executive functioning may serve as a predictor of AD in normative samples (Royall, Palmer, Chiodo, & Polk, 2004). Not only do individuals with AD perform more poorly on these tests of everyday competence, but they also tend to take longer time to complete the test-related tasks and make an increased amount of errors when completing the tasks compared to individuals with MCI (Sacco, Joumier, Darmon, Dechamps, Derreumaux, Lee, Piano, Bordone… & Robert, 2012). The results obtained from the study by Sacco et al. (2012) were derived using scores on the Mini Mental State Examination (MMSE). Scores on the MMSE, another screening measure of cognitive impairment, highly correlate with scores on the DRS, suggesting the generalization of these performance effects to expected scores on the DRS-2 (Folstein, Folstein, & McHugh, 1975; Salmon, Thal, Butters, & Heindel, 1990; Bobholz & Brandt, 1993).

The DRS-2 has been subject to numerous statistically driven studies that aimed to examine the internal structure of the test, or the degree to which the items on DRS-2 actually measure the cognitive constructs that they intend to measure. Consequently, researchers have criticized the DRS-2 for its inconsistency across factor analytic studies. Colantonio, Becker, and Huff (1993) reported a three-factor structure for the DRS, defining the structures as conceptualization, construction, and memory. Consistent findings supporting this three-factor structure have since been reported (Woodard, Salthouse, Godsall, & Green, 1996). Woodard et al. (1996) reported an additional fourth factor that did not meet the researchers inclusion criteria for the identification of factor structures, labeling the non-statistically significant factor as attention/initiation/perseveration. In contrast, additional research identified a two-factor

structure for the measure, in which items can be more appropriately defined solely as verbal and non-verbal components (Kessler, Roth, Kaplan, & Goode, 1994). Kessler et al. (1994) also reported the presence of a high degree of overlap between the Memory and Conceptualization subscales, suggesting that these two subscales are virtually inseparable from one another. Kessler et al. (1994) suggested that the inconsistencies across factor analytic studies might be attributable to the way in which the measure was initially constructed, suggesting that the subscales were created on a strictly conceptual basis. Specifically, researchers have questioned if the subscales lack sufficient empirical support due to the lack of statistical consideration during the initial measurement development process (Kessler et al., 1994). The inconsistency across factor analytic studies could suggest a potential flawed five-factor structure, as was initially determined during the construction of the measure. However, this inconsistency may be better explained by the violation of the independence of scores assumption. The test items fail to be independent from one another due to the arrangement and scoring of the items according to difficulty level, such that a score on one item influences scores on other items. Therefore, exploratory factor analysis may not be an appropriate method for examining the dimensionality of the DRS and the results of these studies should be interpreted with caution.

2.3.1 DRS-2 subscale weighting. The DRS contains a greater number of questions and total points than most general brief cognitive screening tests for dementia, making the test unique from most others. Because of this, breakdown of the subscale scores lend more influential to numerical test outcome, especially when referencing the recommended cut score that implies a level of deficit associated with dementia. As seen

in Table 2, there is an unequal distribution of the number of test items and points allotted to each subscale in relation to the composite. Interestingly, the proportion of subscale test items to the total number of test items on the composite is more evenly distributed than the proportion of number of possible points per subscale to the number of total possible points on the composite. Therefore, it is seen that some subscales were provided with greater statistical weights relative to other subscales in the measure (See Table 2). Although Mattis (1976, 1988) and Jurica, Letten, & Mattis (2001) provided a brief description of the item review process during the development of the measure, no additional information was provided in the testing manual describing the measurement development process. Therefore, it remains unclear as to why certain subscales were weighted more heavily than others when the test was constructed. Furthermore, considering that the DRS-2 was initially created for the purposes of administration to a sample of older adults diagnosed with AD, the current rank ordering of subscale weights appears somewhat a-theoretical. For example, deficit in memory, specifically delayed recall, serves as the most reliable indicator of AD diagnosis (Karantzoulis, & Galvin, 2011; McKhann et al., 2011; Schoenberg & Duff, 2011). Therefore, one may wonder why the Memory subscale did not include the greatest number of questions, or receive the greatest weighting when applying numerical points to each test item. Comparisons of hypothetical subscale performance on the DRS-2 are presented below to illustrate the potential influence of differential subscale weighting on test outcome.

A comparison of hypothetical performance scores on the DRS-2 subscale containing the lowest-weighted score (Construction) and the highest-weighted score (Conceptualization) highlights the effects associated with differentially weighted subscale

scores. For example, the DRS-2 Construction subscale was assigned the lowest weight relative to the other subscales, with a maximum possible subscale score of 6 points (See Table 1). In contrast, the DRS-2 Conceptualization subscale was assigned the greatest weight, with a maximum possible subscale score of 39 points. Interestingly, both the Construction and Conceptualization subscales each contain 6 items. Upon examination of the relative proportions associated with the number of *items* per subscale to the total number of *items* on the entire test, these two subscales provide the same relative contribution to the test. Specifically, the 6 Conceptualization items account for 16% of the total number of items on the DRS-2, and the 6 Construction items account for a separate 16% of the total number of items on the DRS-2. However, this identical pattern in terms of relative contribution of each subscale to the total composite score is eliminated once scores are applied to each of the subscales. These two subscales that initially maintained equivalent weights and contributed equally in terms of number of items per subscale to the number of items for the entire measure become drastically different. Once subscale scores are applied to each of the five subscales, we see that the weight of the relative contribution of the Construction subscale to total score on the composite shifts downward in comparison to its relative contribution associated with the number of subscale items, with a maximum possible subscale score of 6 points. In contrast, the weight of the relative contribution of the Construction subscale score to the total score on the composite shifts upward in comparison to its relative contribution associated with the number of subscale items, with a maximum possible subscale score of 39 points. Now, upon examination of the relative proportions associated with the number of *points* per subscale to the total number of *points* on the entire test (total possible

composite score = 144), these two subscales provide vastly differential relative contributions to the composite score on the measure. With a weight of 39 possible subscale points, the 6 Conceptualization items account for 27% of the 144 total possible points on the overall measure. In contrast, with a weight of 6 possible subscale points, the 6 Construction items account for 4% of the 144 total possible points on the composite.

However, clinicians do not solely refer to the total scores on tests of cognition when making interpretations about patient cognitive performance. Instead, clinicians turn towards more specific components of the test to determine where, or in which cognitive domains, the patient evidenced areas of deficits based on areas of poor performance. Testing data may be misleading, as performance may still be more accurately described as a function of the test. For example, the first out of the 11 tasks on the Initiation/Perseveration subscale of the DRS-2 represents a complex test of category fluency, which has reliably been reported as a specific and sensitive test to detect cognitive deficit in persons with AD (Libon, McMillan, Powers, Massimo, Khan… & Grossman, 2009; Rascovsky & Grossman, 2007). On this item, patients are directed to verbally generate a list of all of the possible examples of items within a specific category. The test technician records the patient's responses over the span of one minute, as this is a time-limited task. Possible scores on this task range from 0-20, with a score of zero representing no correct answers and a score of 20 representing full credit in which all generated responses were scored correctly. Interestingly, in a study conducted by Tombaugh, Kozak, and Rees (1999), cognitive healthy older adults ages 60-69 years generated 17.6 words on a similar test of category fluency, on average. It is expected that

cognitively healthy older adults would earn a perfect score on the DRS-2, as the test was initially created with an intended low floor effect to allow for the assessment of individuals in the severe stage of dementia (Mattis, 1976). However, the research conducted by Tombaugh et al. (1999) suggests that even cognitively healthy older adults would not, on average, earn a perfect score on this particular task. This highlights a point of contention in the development of the measure, as the DRS-2 testing manual does not explain why 20 total possible points were assigned to this task, nor does it provide reasoning for why the generation of 20 items on this list is considered representative of "normal" functioning. Based on the current distribution of points across tasks from this subscale, it appears as though a loss of points on this one task may result in a misrepresentation of cognitive deficit.

For example, if a 65 year-old patient received a score of 17/20 on this one task, a score approximately representative of the average performance of a cognitively healthy older adult of this age, but received full credit on all subsequent subscale tasks, the patient would earn a total subscale score of 34/37. In comparison to the performance of individuals of the same age, the MOANS scaled scores indicate that a subscale score of 34/37 falls in the 11-18th percentile range. Instead, it may make more sense for the task to be assigned a value of 17 total possible points, as opposed to 20, to more accurately reflect evidence of cognitive deficit when scores fall below the average score of age-matched cognitively healthy comparisons. It remains unclear as to why this task is scored out of a total of twenty possible points, as this could presumably result in a loss of points from healthy older adults without dementia. Therefore, this particular task renders the AD patient increasingly susceptible to earning what may be considered unjustifiably

low scores on this task. This highlights the need for additional consideration to be given to test development efforts when determining the relevant importance of tasks and subscales within a measure of cognition.

III. Current Study

Statistical and empirical support should be considered in tandem during measurement construction in order to further improve upon the reliability and validity of cognitive tests. Even though test developers provide strong support for some forms of reliability and validity evidence (e.g., interrater reliability, internal consistency, test-retest reliability, validation of test questions by several content experts), it appears that scholarship has devoted less attention to statistical components vital to the measurement construction process. The DRS-2 testing manual, for example, does not provide statistical explanation regarding the distribution of points to each subscale. Conceptually, the amount of points assigned to each subscale do not appear to align with the empirical research regarding the diagnosis of either all-cause dementia or AD dementia, specifically. Specifically, research has reliably identified performance on delayed recall tasks as the best indicator of the presence of AD, as neuropathological effects in AD are first identified in the medial temporal lobes, the area of the brain responsible for memory consolidation (Karantzoulis & Galvin, 2011; McKhann et al., 2011; Schoenberg & Duff, 2011). Therefore, the existing literature suggests that the memory subscale maintains the greatest importance in the prediction of cognitive deficit associated with AD. However, that is not reflected in the assignment of points to each of the DRS-2 subscales (See Table 1). Although the DRS-2 is administered to track cognitive decline in individuals with various types of dementia, the original DRS testing manual stated that the measure was

initially developed to assess the cognitive functioning of dementia patients diagnosed with Alzheimer's disease. Therefore, considering this information in conjunction with the characteristics of the original sample from which the test was first normed, it appears that the Memory subscale warrants the greatest importance relative to the other subscales. This is not the case, however, as the Memory subscale maintains the lowest relative contribution to the total number of questions and the total number of possible points on the entire measure out of all five subscales with respect to the number of items it contains (5 items, 13% of total test items included in the entire measure). With no explanation provided in the testing manual to support the specific ways in which this measure and its subscales were constructed, it appears that arbitrary weighting procedures may have been inadvertently applied during the development of the DRS-2. This uncertainty warrants further examination into the relative importance of each of the DRS-2 subscales to determine sources of statistical support for the relative importance implied by the corresponding points assigned.

Taken together, literature suggests a differential effect of the weighting of subscale scores on measurement reliability and validity. Specifically, research suggests that the accuracy of cut scores may highly relate to the internal structure of a measure (Baldwin, 2015; Burt, 1950; Govindarajulu, 1988; Meehl & Rose, 1955; Perloff & Persons, 1988; Thompson, 1940; Wang & Stanley, 1970). Although cut scores are not utilized solely for clinical decision-making, the proposed cut scores associated with the DRS-2 may vary across the clinical population in which the test is administered, due to the fact that the empirically supported differential trajectory of cognitive decline is not reflected appropriately for all clinical populations in the relative importance of the DRS-2

subscale scores. Therefore, there is a need for the evaluation of statistical support for the weighting methods applied to both new and pre-existing neuropsychological measures in order to ensure that these measures are constructed optimally to best allow for the early detection of dementia. Consideration should be given to cases in which support for alternative weighting methods are conceptually and statistically supported, as improvements to these measures may increase their ability to pick up subtle changes associated with dementia, therefore allowing for more increased correct classification rates of the disease when utilizing these screening measures.

3.1 Research Questions

1. What is the ordering of relative importance of the five DRS-2 subscales to predicting dementia diagnosis?
2. What are the effects of utilizing a logit-weighting subscale algorithm on DRS-2 sensitivity and specificity compared to the sensitivity and specificity of the original version of the test?

3.2 Hypotheses

Prior research suggests that that memory impairment appears first in the course of progressive cognitive decline associated with AD dementia, with impaired delayed recall serving as the strongest predictor of the presence of AD dementia (Schoenberg & Duff, 2011). The clinical manifestation of episodic memory impairment arises as the result of the degeneration of the medial temporal lobe structures, including the entorhinal cortex and the hippocampus (Karantzoulis & Galvin, 2011). Impaired language abilities are also one of the associated impairments evidenced early on in the course of AD dementia (Karantzoulis & Galvin, 2011). Specifically, patients with AD display marked

impairments in semantic fluency compared to healthy older adults of the same age (Lezak et al., 2012). In addition, research suggests visuospatial deficits are also prominently noted early on in the typical course of AD (Possin, 2010; Rizzo, Anderson, Dawson, & Nawrot, 2000). These deficits are often apparent on neuropsychological tests in which the patient is required to complete a figure-drawing task without error. Although executive functions are, in most cases, spared in the earliest stages of AD, dysfunction in this cognitive domain typically surfaces during the moderate stage of AD and continues to worsen with progression into the severe stage (Baudic, Barba, Thibauder, Smagghe, Remy, & Traykov, 2006). These deficits most commonly present on tests of cognitive ability as difficulties in abstract reasoning, with signs of decreased ability to complete tasks, follow directions, and transition from one task to another arising more prominently in the moderate stage of AD (Crowell, Luis, Vanderploeg, Schinka, & Mullan, 2002; Collette, Amieva, & Adams, 2007). Research suggests that attention remains relatively preserved in patients with AD, and if often one of the last cognitive processes affected by the disease (Calderon, Perry, Erzinclioglu, Berrios, Dening, & Hodges, 2001). Taken together, I hypothesized that the logistic regression analysis utilized to investigate my first research question would reveal the following rank ordering, from greatest to least importance, of the five DRS-2 subscales: Memory, Initiation/Perseveration, Conceptualization, Construction, Attention. With respect to my second research question, I hypothesized that the new scoring algorithm generated from the standardized logistic regression logit weights would increase the sensitivity, and either decrease or maintain the same level of specificity of the DRS-2 compared to the sensitivity and specificity rates utilizing the original version of the test in this sample.

IV. Methodology

This study utilized archival data from the Adult Neurology Clinic at the University of Virginia Health System located in Charlottesville, Virginia to obtain the total sample, which included 113 older adults ages 65 and older (55 AD, 58 MCI). These patient groups were selected for use in the current analyses because these were the groups in which DRS-2 data were available within the archival database in which data were obtained. Table 3 presents a summary of demographic information for both participant groups.

A Data Use Agreement was drafted between James Madison University and the University of Virginia to allow for the use of this limited archival dataset. No identifying information was contained in the archival database. The following variables were included in the database: randomly generated and non-linked case ID, group, age, sex, race, education level, and DRS raw subscale and total scores. No missing data were identified within the dataset. The UVA Institutional Review Board approved the creation of a patient database for research purposes prior to the initiation of data collection. Data in this dataset were obtained during the participants' initial comprehensive flexible neuropsychological battery administration that took place during the clinical diagnostic workup process. A trained psychometrician or clinical neuropsychology post-doctoral fellow administered the DRS-2 in the English language. Rater agreement information was not recorded within the archival database, and therefore interrater reliability was unable to be calculated. The neuropsychological evaluations administered utilized a flexible battery approach to obtain the quantitative data for the cognitive assessment portion of the evaluation. All participants in this group were clinically diagnosed with

either AD dementia or MCI suspected to be due to AD by a licensed clinical neuropsychologist. Diagnoses of suspected AD etiology were made based on all available information at the time of testing, which often included a neurology workup, labs, and imaging, in addition to the results from the comprehensive neuropsychological evaluation. Diagnoses were made in accordance with the diagnostic criteria presented in the McKhann et al. (2011) and Petersen et al. (2014) diagnostic criteria guidelines for AD and MCI, respectively. All analyses were conducted using IBM SPSS Statistics Version 25 (IBM Corp., 2017).

The new scoring algorithm for the DRS-2 was constructed utilizing the standardized beta weights generated using hierarchical single predictor entry logistic regression. Variables were entered independently into the logistic regression equation in order of predictor importance as suggested by the results of empirical studies outlined in the literature review.

V. Results

5.1 Preliminary Analyses

Prior to analyses, all data were examined via visual inspection and frequency and descriptive analyses in SPSS for accuracy of data entry, missing values, outliers and fit between their distributions and the assumptions associated with logistic regression analysis. Variables were examined separately for the 55 AD participants and the 58 MCI participants. There were no missing data identified in the dataset.

A preliminary series of independent samples *t*-tests were conducted to assess whether there were significant mean differences in demographic values between the AD and MCI participant groups. On average, group ages differed by 2.190 years, which was

a statistically significant difference at an alpha level of .05, $t(111) = 2.060$, $p = .042$, CI: .083 to 4.296. AD and MCI participants differed by an average of .032 years of education, which was not statistically significant at an alpha level of .05, $t(111) = .053$, $p = .958$, CI: -1.162 to 1.226. A summary of the means and standard deviations for age and education level associated with each of the two groups is presented in Table 2.

Additional independent samples t-tests were conducted to identify mean differences between groups on each of the continuous variables included in the logistic regression analysis to foreshadow which variables would likely provide the greatest relative utility once entered into the logistic regression equation. Of the five subscales, AD and MCI patients statistically significantly differed on Attention [$t(111) = -3.173$, $p = .002$, CI: -3.123 to -.722], Initiation/Perseveration [$t(111) = -5.396$, $p < .001$, CI: -6.734 to -3.117], Conceptualization [$t(111) = -3.647$, $p < .001$, CI: -3.680 to -1.072], and Memory [$t(111) = -9.991$, $p < .001$, CI: -6.835 to -4.573]. Groups did not statistically significantly differ on Construction [$t(111) = -1.844$, $p = .068$, CI: -.510 to .018]. See Table 4 for a summary of DRS-2 subscale mean scores and standard deviations. Comparisons of average test scores across the five subscales foreshadowed that the four subscales in which the two groups differed (Attention, Initiation/Perseveration, Conceptualization, and Memory) would potentially provide the greater unique utility in discriminating between AD and MCI patients, relative to the Construction subscale. Comparisons of unique predictor utility were further tested in the logistic regression analysis via comparison of standardized predictor weights. Results from this analysis are reported below.

5.1.1 Assumptions. All assumptions associated with logistic regression were tested. First, logistic regression assumes the correct specification of the model, such that all IVs specified in the hypothesis are included in the model (Cohen, Cohen, West, & Aiken, 2003). Visual inspection of the scatterplots of the residuals for each IV included in the regression model did not reveal the presence of substantial systematic variation. Therefore, there did not appear to be any support for a violation of this first assumption (Cohen et al., 2003).

Second, logistic regression assumes the independence of errors, such that cases included in the dataset are unrelated. Errors are assumed to be independent in between-subjects research designs. Therefore, given that the current study utilized a between-subjects research design, analyses were conducted without concern of a violation of this assumption.

Third, logistic regression assumes no measurement error in the IVs. Violations of this assumption result in biased parameter estimates and standard errors, increased type 1 error rate, and attenuated effect sizes associated with individual estimates (Cohen et al., 2003). Coefficient alpha was unable to be computed as a measure of internal reliability for this measure, as item-level information was not recoded in the archival database utilized for this study. Similarly, interrater reliability was unable to be calculated, as no information regarding interrater agreement was collected and recorded in the database. However, the DRS-2 was administered to all patients and scored by a psychometrician or a clinical neuropsychology post-doctoral fellow, both of whom completed the same UVA neuropsychological testing administration training. This suggests that appropriate actions were taken to ensure sufficient interrater reliability and therefore decreased the possibility

that rater disagreement posed a notable threat to reliability. Error committed during the data entry process must also be considered when examining for violations of this assumption. Specifically, data entry error can create problems that closely resemble issues associated with measurement error and subsequently be mistakenly interpreted as such. All data included in the archival database were initially obtained via medical record review and through an additional review of paper copies of patient medical files. This process allowed for the examination of consistency between both sources of information, reducing the possibility for data entry error. I proceeded with analyses recognizing my inability to calculate reliability estimates as a limitation of the study.

5.1.2 Practical issues. Multicollinearity among the IVS, outliers, and complete separation of the data between two participant groups may cause several issues when conducting logistic regression. Specifically, high correlations among the IVs in the model result in inflated standard errors associated with the parameter estimates (Tabachnik & Fidell, 2013). Pearson correlation coefficients were computed to examine the bivariate relationships among the five predictors in the model and to assess for potential multicollinearity (see Table 5). Higher education level was statistically significantly related to higher scores on Initiation/Perseveration ($p < .05$) and Conceptualization ($p < .01$). Scores on Attention were positively related to scores on Initiation/Perseveration ($p < .01$) and scores on Conceptualization were positively related to scores on Memory ($p < .01$). Initiation/Perseveration scores also positively correlated with Memory ($p < .01$) and Conceptualization ($p < .05$) scores. None of the correlation coefficients were greater than or equal to a value of .70, suggesting that there was not a high degree of overlap among the predictors (Cohen, 1988; Tabachnick & Fidell,

2013). Therefore, I proceeded with the logistic regression analysis without any major concern regarding issues of multicollinearity; however, the overlap among the predictors was considered when interpreting results of the logistic regression analyses.

The presence of outliers in logistic regression can result in biased parameter estimates and corresponding standard errors. Outliers were examined separately for AD participants and MCI participants. No outliers in the dataset were identified via visual inspection of the residual scatterplots, box plots, or histograms for each of the five predictors.

The presence of sparseness among the data can result in a continuous inflation of the parameter estimates and corresponding standard errors (Tabachnik & Fidell, 2013). It can also cause convergence failure of the model. Sparseness was examined by comparing data from patients diagnosed with AD to data from patients diagnosed with MCI for each of the five DRS-2 subscales. No groupings of empty cells or low cell counts were identified among the data, suggesting data were not sparse.

5.2 Primary Analyses

Logistic regression was conducted to investigate the relative contribution of the set of DRS-2 subscale scores to predict cognitive deficit associated with a diagnosis of dementia. A logistic regression model including the five subscales as predictors was fitted to the data to test the hypothesis that the five tests included in the model would differentially contribute to patient diagnosis.

Logistic regression results showed that,

Predicted logit of DIAGNOSIS = -28.192 + (.171)*ATT + (.186)*I/P + (.157)*CONCEP + (.495)*MEM + (.408)*CONST. **(1)**

The model containing all five DRS-2 subscales (-2LL = 72.456) statistically significantly improved the prediction of group membership (AD, MCI) compared to a model containing no predictors (D_{Null} = 156.57) , χ^2 (5, N = 113) = 84.12, p <.001. The amount of null deviance left unexplained by the five-predictor model was not statistically significantly different from the amount of deviance that would be left over due to chance [-2LL (72.456) < $\chi^2_{\mathbf{crit}}$ (113.145)]. Therefore, I concluded that the model containing all five DRS-2 subscales adequately predicted group membership.

Approximately 51.5% of the null deviance is accounted for by Attention, Initiation/Perseveration, Conceptualization, Memory, and Construction DRS-2 test scores (R^2_L = .515). Two additional R^2 analog indices indicate that the five-predictor model accounts for 52.5% (Cox and Snell R^2 = .525) and 70% (Nagelkerke R^2 = .700) of the null deviance in the criterion.

Memory was a significant predictor of AD diagnosis, controlling for all other predictors in the model [Wald χ^2 (1, N=113) = 24.383, p < .001]. Specifically, for every unit increase in Memory, the odds of AD diagnosis increased by a factor of 1.641 [CI: 1.348 to 1.998]. Initiation/Perseveration was also a significant predictor of AD diagnosis, controlling for all other predictors in the model [Wald χ^2 (1, N=113) = 6.086, p = .014]. For every unit increase in Initiation/Perseveration, the odds of AD diagnosis increased by a factor of 1.204 [CI: 1.039 to 1.396]. See Table 6 for a summary of the predicted log odds and odds ratios associated with each of the five predictors included in the logistic regression model.

Comparisons were made between actual group membership and predicted group membership using the five-predictor model. Overall, 85.8% of the participants were correctly classified with a diagnosis of dementia using the model containing all five predictors. Specifically, 48 out of 55 participants that were actually diagnosed with dementia due to AD were correctly classified as belong to the AD dementia group (87.3%). In contrast, 49 out of 58 participants that were actually diagnosed with MCI were correctly classified as belonging to the MCI group (85.8%).

Memory [b = .495, CI: .299. to .692, OR = 1.641, CI: 1.348 to 1.998, Wald χ^2 (1, N=113) = 24.383, p < .001, Likelihood Ratio χ^2 (1, N=113) = 64.739, p < .001] and Initiation/Perseveration [b = .186, CI: .038 to .334, OR = 1.204, CI: 1.039 to 1.396, Wald χ^2 (1, N=113) = 6.086, p = .014, Likelihood Ratio χ^2 (1, N=113) = 8.578, p = .003] each statistically significantly accounted for the outcome, when controlling for all of the other predictors in the model. The Wald tests were non-significant for the three other predictors in the model (p > .05).

Each of the predictors in the model was converted to standardized z-scores and the standardized scores were entered into a second logistic regression equation to allow for comparisons regarding relative utility to be made across all five predictors. The DRS-2 Memory subscale had the greatest predictive utility to detecting cognitive deficit associated with dementia (β = 2.062, OR = 7.863) given all of the other predictors in the model. Initiation/Perseveration provided the second greatest predictive utility given the other variables in the model (β = 1.008, OR = 2.741). Out of all five tests, Conceptualization provided the third highest predictive utility of clinical diagnosis (β = .576, OR = 1.779). Attention (β = .573, OR = .1.774) and Construction (β = .292, OR =

1.339) provided the fourth and fifth ranked predictive utility, given all of the other predictors in the model. The standardized log odds for each of the five predictors were utilized to generate the new subscale weights for the new scoring algorithm of the DRS-2 subscales.

5.3 Secondary Analyses

To answer my second research question, I examined the predictive validity of the DRS-2 cut score using the new logit-weighting algorithm. Comparisons were then made between the original DRS-2 scores and the newly computed DRS-2 scores generated by the statistical weighting algorithm. To do this, I utilized the recommended cut score of 123/144 to identify cases representative of true positives, false positives, true negatives, and false negatives for the original DRS-2 raw total scores. I calculated sensitivity and specificity for the original form of the test utilizing this data displayed in Figure 1. In the current study, sensitivity represents the percentage of participants with dementia correctly identified by the DRS-2 as having dementia, whereas specificity represents the percentage of participants without dementia correctly identified as not having dementia (Lezak et al., 2012). In this sample, the original form of the DRS-2 maintained 78.18% sensitivity and 86.21% specificity. These rates were later compared with the sensitivity and specificity rates generated utilizing scores derived from the new weighting, or scoring, algorithm. The steps outlined below provide a detailed description of how the newly weighted DRS-2 subscale scores were derived.

The initial weighting of the subscales is presented in Table 2. A comparative display of the new weights generated by the logistic regression analysis is presented in Table 7. The log odds values obtained from the unstandardized regression equation

served as the new weights for the re-weighted subscale scores, replacing the original weighting values shown in the center column of Table 2. New scores were computed by multiplying the total number of questions per subscale by the associated unstandardized logit for each subscale. As shown in the bottom row of Table 7, applying these weights generated a new total DRS-2 score of 31.19. However, in order for sensitivity analytic comparisons to be made between the two different scoring methods, these re-weighted values were then put back on the same metric, with a total possible DRS-2 score of 144 utilizing the equations displayed in Table 8. Once the new total possible subscale scores were computed, original raw subscale scores were then re-weighted and converted to new, logit-weighted scores on the same metric. The equations are presented below.

To compute the new metric-consistent and logit-weighted subscale scores,

NewMEMScore = (OriginalMEMRawScore *NewMEMTotalPossibleScore) /OriginalMEMTotalPossibleScore **(2)**

NewIPScore = (OriginalIPRawScore *NewIPTotalPossibleScore) /OriginalIPTotalPossibleScore **(3)**

NewCONCScore = (OriginalCONCRawScore *NewCONCTotalPossibleScore) /OriginalCONCTotalPossibleScore **(4)**

NewATTScore = (OriginalATTRawScore *NewATTTotalPossibleScore) /OriginalATTTotalPossibleScore **(5)**

NewCONSTScore = (OriginalCONSTRawScore *NewCONSTTotalPossibleScore) /OriginalCONSTTotalPossibleScore **(6)**

I then calculated the new logit-weighted total raw score for each participant in the sample by adding each of the new subscale raw logit-scores. Through these calculations, the logit-weighted total possible score was then consistent with the original DRS-2 total possible score (i.e., total 144 possible points). Utilizing the newly computed logit-

weighted total scores for the participants, I calculated a second, comparative sensitivity analysis to answer my second research question (i.e., Will statistically-weighted subscale scores improve the sensitivity of the DRS-2?). Again, I counted the total number of true positive, false positive, true negative, and false negative cases. Figure 2 presents a summary of these values. Utilizing the same recommended 123 cut score as before, I calculated sensitivity and specificity rates for the new logit-weighted total DRS-2 scores. In this sample, the logit-weighted scores of the DRS-2 maintained 90.90% sensitivity and 74.14% specificity. This represented a 12.72% increase in sensitivity and a 12.07% decrease in specificity. Despite the drop in specificity, the increased sensitivity of the measure by utilizing the statistically derived subscale weights provided support for my second hypothesis that the logit-weighted scores would increase the sensitivity of the measure to detecting cognitive deficit associated with dementia.

VI: Discussion

6.1 Primary Hypotheses

The first goal of this study was to examine the relative contribution of each of the five DRS-2 subscales to predicting AD diagnosis. Partial support for my first hypothesis was revealed from the logistic regression analysis. Results revealed the following rank ordering, from greatest to least importance, of the DRS-2 subscales to predicting AD diagnosis: Memory, Initiation/Perseveration, Conceptualization, Attention, and Construction. These results differed from my initial hypothesis such that Attention ranked as the fourth subscale of greatest importance, as opposed to my predicted fifth and least contributing subscale in terms of relative unique utility per my initial hypothesis. It is possible that the items within any of the five subscales tapped into the assessment of

additional cognitive abilities, above and beyond the construct intended for assessment within the particular subscale. For example, the Attention subscale may have tapped into other areas of cognition, such as motor coordination, language abilities, and working memory. With language abilities being one of the earlier cognitive functions affected by Alzheimer's disease early on in the diagnosis, it is possible that this could explain the statistical rank ordering of this subscale as the fourth greatest contributor to the model containing all five DRS-2 subscales. According to the description provided in the DRS-2 professional manual, "both auditory-visual and verbal-nonverbal tasks of attention are presented" in the Attention subscale (Jurica et al., 2001). However, one of the eight tasks within this section is digit span backward, which requires the test-taker to hold a sequence of numbers in his or her head before repeating them in reverse order to the person administering and scoring the test. Digit span backward is generally understood as a test of working memory, as it requires active mental manipulation of information (Emrani, Libon, Lamar, Price, Jefferson, Gifford… & Au, 2018). In a recent study conducted by Emrani et al. (2018), the authors report that MCI patients displayed an attenuation of a recency effect when administered the digit span backward test, compared to normal controls without evidence of cognitive deficit. This suggests that patients in the AD group were likely scoring well below the patients in the MCI group on this particular task within the subscale, due to the continued neurodegenerative effects associated with Alzheimer's disease progression. Therefore, this may have introduced an additional amount of variability into the subscale scores, which may have resulted in the rank ordering of this predictor as the fourth greatest contributing model, relative to the other predictors in the model.

6.2 Secondary Hypotheses

The second goal of this study was to compare the sensitivity and specificity rates of the logit-weighted DRS-2 total scores to the sensitivity and specificity rates of the original DRS-2 total scores, when utilizing the recommended 123/144 cut score as a criterion. Analyses revealed that the logit-weighted DRS-2 scoring method yielded a higher sensitivity compared to the original version of the DRS-2, such that the sensitivity of the measure increased by 12.72%. In contrast, analyses revealed that the logit-weighted DRS-2 scores yielded a lower specificity compared to the original DRS-2 scores, such that specificity rates decreased by a 12.07% using the logit-weighted scores. The decrease in specificity may be attributable to the increased weighting of subscales testing areas of cognition that are reliably reported to distinguish cognitive deficit associated with a diagnosis of AD dementia and individuals not diagnosed with dementia, such as Memory. Increasing the weighting of the subscales of highest importance likely dropped the scores on these subscales for the initial false negative cases, resulting in re-weighted scores below 123, and therefore accurately re-classifying these cases as true positives. The decrease in specificity may be due to the increased range of the distribution of total points possible across subscales. In the original form of the DRS-2, the total possible points across subscales ranged from 6 to 39 points. In the re-weighted version of the DRS-2, the total possible points across subscales ranged from approximately 8 to 51. The difference between the total possible subscale points between the highest and lowest weighted subscales differed by 33 points in original DRS-2. In comparison, in the logit-weighted DRS-2, this difference increased to 43 points. The range of these differences may have affected the calculated specificity of the two scoring

methods of comparison, as narrowness generally lends itself to increased ability for unique identification of cases of interest.

6.3 Limitations

Several statistical limitations must be considered when interpreting the results from this study. The logistic regression analysis was underpowered due to small group sample sizes, which likely upwardly biased the produced parameter estimates and inflated the associated standard errors (Tabachnick & Fidell, 2013). Conventional recommendations indicate a minimum sample of 60 participants within each of the two groups (AD, MCI), resulting in a total sample size of 120 participants. This determination of the minimum sample size was based on the recommendations proposed by Peduzzi, Concato, Kemper, Holdord, and Feinstein (1996) that logistic regression analyses should contain a minimum of ten observations per parameter estimate. The logistic regression equation for the current study included six total parameter estimates: one estimate associated with the intercept in the equation and one estimate per each of the five DRS-2 subscales (See Formula 1).

Additionally, the standard errors associated with the logistic regression coefficients may be biased due to the moderate degree of shared variability among all five of the predictors in the model, as indicated by the values of the intercorrelations among the predictors presented in Table 5. For example, moderate degrees of overlap were evident among the following pairings of predictors: Memory and Initiation/Perseveration, Memory and Conceptualization, Conceptualization and Education, Initiation/Perseveration and Education, Initiation/Perseveration and Attention, and Initiation/Perseveration and Conceptualization. These findings further support the

results presented by Marcopulos et al. (1997) highlighting the effect of education on neuropsychological test scores, as displayed by the positive relationship between four out of the five DRS-2 subscale predictors and education level. The exception to this finding was the Attention subscale, which revealed a small negative correlation with education level. Although likely trivial in nature, it is possible that this, too, could be due to the possibility that the items within the Attention subscale tap into additional cognitive abilities, as previously mentioned.

It is also important to note that, although the correlations were small, each of the predictors, with the exception of Conceptualization, was negatively correlated with age. This is finding is consistent with the literature supporting a decrease in cognitive function with increasing age. Additionally, it is important to note that the AD group participants were slightly older than the MCI group participants. Taken together with the results of the primary and secondary hypotheses, this provides support for decreased functioning across cognitive domains for AD patients. The DRS-2 testing manual describes that the Conceptualization subscale is designed to assess "the ability to abstract, that is, to induce similarities and detect differences among verbal and visual stimuli" (Jurica et al., 2001). Due to the combination of verbal and non-verbal abilities, as well as concrete and abstract reasoning abilities that are assessed within the Conceptualization subscale, it is possible that the items better represent several different, unique constructs, as opposed to one, larger construct that was labeled "Conceptualization." This may have decreased the internal consistency of this particular subscale due to potential variability in what is being assessed, and therefore may explain the small negative correlation with age reported in this study.

Furthermore, the overlap among the predictors may have attenuated the effect of the unique contribution of the overlapping predictors to the outcome, given the other predictors in the model. Ultimately, this could have resulted in the non-significance of individual predictors as well as attenuated effect sizes associated with the predictors that shared a greater amount of variance with other predictors in the model (Cohen et al., 2003).

Another limitation of this study is the lack of collection of item-level information in the archival database used for the analyses. Without item-level information, I was unable to examine reliability of the DRS-2 in the current sample, and therefore was unable to examine for the presence of measurement error. Measurement error in the predictors is associated with downward bias of the associated individual regression coefficients (Tabachnick & Fidell, 2013). Measurement error is also associated with an inflation of the standard errors associated with the individual regression coefficients, as well as inflation of the standard errors of the outcome (Tabachnick & Fidell, 2013).

Practical limitations must also be considered when interpreting these results. It is important to note that the DRS-2 was administered within the comprehensive evaluation that was part of the initial diagnostic workup process for all participants. Therefore, the DRS-2 testing information was available and utilized in conjunction with other cognitive tests administered during the flexible comprehensive test battery as well as with all other available information at the time of testing to make a clinical diagnosis of MCI or dementia. Although there were many other factors considered during the diagnostic workup for these patients, the use of the DRS-2 introduces a source of bias due to the

presence of some degree of circular reasoning into the results of the study, as the diagnoses were not completely independent of the test scores used for analyses.

A second notable limitation of this study pertains to the cut score used and the use of MCI participants as a control, or comparison, group. The use of the 123/144 recommended cut score to classify participants may not have been appropriate, as this cut score was created utilizing the original weights applied to the DRS-2 subscale scores. Therefore, it is possible that a different cut score may be deemed more appropriate for use if a normative study were to implement a logit-based weighting algorithm to derive new DRS-2 scores. Additionally, it is important to note that this original 123/144 cut score was not developed to distinguish between levels of cognitive deficit associated with dementia and MCI. Instead, it was developed to distinguish cognitive deficit indicative of dementia relative to a normal controls, or persons with no evidence of abnormal cognitive impairment. Therefore, the use of this cut score in the current study may not have been appropriate, and may have introduced bias into the sensitivity/specificity comparisons of the two scoring methods. Furthermore, the unstandardized beta weights generated from logistic regression equation in which the MCI group served as the non-case group will likely differ from the unstandardized beta weights that would be generated from a logistic regression equation in which a normative sample served as the non-case group. Future research should be conducted to investigate these differences and the effect of statistical weighting procedures using a true normative comparison group.

Additionally, I was unable to control for individual differences on a number of factors that may have influenced cognitive performance. For example, the archival database utilized for this study did not contain information regarding comorbid medical

conditions. It is possible that DRS-2 scores may have been influenced by the presence of another medical condition, such as depression or anxiety. Therefore, it is possible that low scores did not solely reflect deficits associated with AD or MCI. Specifically, poor cognitive performance may have been exacerbated by the presence of one or more coexisting medical conditions, or instead may have been better explained by a comorbidity altogether.

The inability to control testing conditions across participants poses another limitation to this study. Specifically, it is highly unlikely that the same person administered and scored the DRS-2 to all participants included in this study. Additionally, the DRS-2 testing data was obtained for each participant during his or her initial diagnostic work up at the clinic via administration of a flexible neuropsychological battery. Therefore, the order of the administration of the tests included within each unique battery may have differed, meaning that the DRS-2 may have been administered to each participant at different times within the evaluation. This may have influenced participant performance on the DRS-2 due to potential differences in the levels of testing fatigue among participants who may have been administered the test in the beginning of the battery in comparison to participants who may have been administered the DRS-2 at the end of the battery. It may have been possible that participants who were administered the test towards the end of the battery performed more poorly than those who were administered at the beginning of the battery, if the ordering of testing administration did, in fact, differ. Additionally, participants who may have been administered the DRS-2 at the end of the battery may have performed more poorly on the test than they would have if they were administered the test at the beginning of the battery.

Another limitation of the study is the use of clinical diagnosis to classify participant group membership. Currently, the only way to confirm a diagnosis of AD is through post-mortem biological analysis. Reports of AD biomarkers were not included in the dataset and I was therefore unable to determine that biomarkers were utilized in conjunction with the results of the comprehensive neuropsychological evaluation to inform clinical diagnosis. Therefore, this remains a limitation of this study.

6.4 Potential Implications and Recommendations for Future Research

The results from this study contribute to the current literature seeking to identify reliable differences in impairment across cognitive domains that best discriminate patients diagnosed with AD and MCI. Specifically, tasks within the DRS-2 Memory and Initiation/Perseveration subscales statistically significantly discriminated between AD and MCI groups. This is consistent with a diagnosis of dementia, which is characterized by insidious onset and gradual decline objectively evident in two or more cognitive domains (McKhann et al., 2011). Furthermore, prior research suggests that cognitive decline is seen in these areas much earlier on in the disease process (Karantzoulis & Galvin, 2011; Lezak et al., 2012; Schoenberg & Duff, 2011;). Therefore, performance on tasks assessing these areas of cognition often progressively declines as a result. On the DRS-2, was evident by the binary statistically significantly lower performance of the AD group in comparison to the MCI group on the Memory and Initiation/Perseveration subscales, such that the differences then contributed to the model at the multivariate level. However, additional research allowing for analysis at the item-level analysis is needed in order to investigate the utility of specific tasks within the discriminating subscales to separate the two groups. For example, delayed recall is the most reliable

indicator of cognitive deficit associated with AD dementia (Karantzoulis, & Galvin, 2011; McKhann et al., 2011; Schoenberg & Duff, 2011). Therefore, it would be expected that this particular task, which is housed within the Memory subscale of the DRS-2, would hold much more weight in discriminating persons with AD dementia from persons without AD dementia. Unfortunately, item-level information was not recorded in the archival database and was therefore unable to be analyzed. Future researchers should consider expanding beyond this limitation of the current study and continue to explore the utility of individual tasks within brief screening measures of cognition to discriminating between AD dementia and non-AD dementia participant groups.

Additionally, results support previous research highlighting the importance of weighting procedures, therefore warranting increased consideration from test developers and clinicians when creating new measures, modifying pre-existing measures, and selecting measures for use across various clinical populations. These results highlight that the use of alternative weights may improve classification rates, which warrants additional consideration in the research and clinical fields. Additional research is needed to see if the results of the current study replicate across diverse demographic populations as well as in bigger sample sizes of both AD and MCI participants and AD and normal controls.

Future research is needed to investigate the utility of statistically driven weighting methods in other clinical samples as well to test for consistency in the results reported in the current study across clinical populations. It is possible that differential scoring algorithms for different clinical populations may further increase the utility of screening tests. Optimally weighted cognitive screening measures may better, or more accurately

inform physicians when making decisions regarding patient treatment. It is not to be implied that screening tests should replace the administration of comprehensive neuropsychological batteries, as screening tests are utilized for completely different purposes than comprehensive evaluations. However, the results of this study do suggest that there is room to further improve upon pre-existing cognitive screening measures and test development procedures in the field of neuropsychology.

Directing efforts towards increasing the sensitivity of these brief cognitive screening instruments also warrants consideration of the use of appropriate cut scores. Many cut scores are derived at the time of measurement construction, or only shortly after. For example, the most frequently referenced cut score for the DRS_2 was created in the 1970s. Moving forward, a re-evaluation of standard setting procedures used to determine these cut scores may be necessary. Specifically, the results of this study highlight the differential effects of subscale scores on patient classification utilizing the 123 recommended cut score. Despite the limitation of using this cut score for comparison of the new scoring algorithm, although it was derived from the original scoring method, this study highlights the need for consideration of standard setting within specific domains of cognition instead of across all cognitive domains being assessed. More specifically, it may make more sense to set standards within each appropriately defined cognitive domain (i.e., memory) and to make domain-specific cut score comparisons to provide physicians administering these brief screening measures of cognition with more specific information to make more informed referral decisions. Cross validation studies are needed in future research in order to examine the

generalizability of standard setting and content-related validity evidence across independent samples.

The results from this study highlight the need for statistical consideration to be given during and reported as a part of the measurement development process. This study provides an example of the ways in which alterations to statistical weights of sub-components of a measure can affect the utility of the test. Although this study presents a new statistically generated scoring algorithm for the DRS-2, at this time, I do not recommend the use of population-specific regression-based scoring algorithms in clinical practice. The current study only investigated the effects of utilizing a regression-based scoring algorithm on one particular clinical population of interest. Therefore, these results would be different across different clinical populations, resulting in the generation of different weights across these different populations. It is unclear to what extent these weights would differ across persons with different types of dementias and it is also unclear if similar results would be found using a different kind of statistical weighting technique other than logistic regression. However, this current study highlights the need for additional research into the aforementioned possibilities. If future research supports the results of this current study, the development of alternative scoring methods may have important clinical implications, as the use of new methods may further improve detection of cognitive deficit associated with abnormal aging and decline. However, this would first require the cross-validation of these findings in order to better understand the generalizability of these findings across many independent samples.

Instead, at this time, a more appropriate recommendation is made for test developers to consider these factors when designing new instruments and to consider the

effects of these decisions on measurement sensitivity and specificity rates. For example, sensitivity is frequently valued in brief cognitive screening tests from a clinical perspective, as the pros of a false positive result (i.e., obtaining a baseline full neuropsychological assessment for future comparative purposes) outweigh the cons of obtaining a false positive result (i.e., utilization of time and financial resources). In other words, these measures should aim to minimize Type II error, as the intent of use of these brief cognitive screening measures is ultimately to increase the identification of deficit associated with abnormal aging (i.e., dementia). Consequently, this results in an increase in Type I error for these measures, although this type of error is of much less concerning in the brief screening setting, due to the push to increase referrals for a full diagnostic workup via comprehensive neuropsychological testing, if deemed appropriate. It is important to consider the detrimental effects of committing a Type I error when diagnosing dementia, as persons diagnosed with dementia are at increased risk for suicide (Draper, Peisah, Snowdon, & Brodaty, 2010). Research suggests that this risk increases for patients diagnosed with dementia during hospitalization, which is a common location in which these brief cognitive screening measures are administered (Erlangsen, Zarit, & Conwell, 2008). Although brief cognitive screening measures are not used independently to diagnose dementia, it is important that professionals administering these tests and making referral decisions for diagnostic workups clearly communicate the implications of the results of brief cognitive testing to the patient. Furthermore, professionals must remain aware of the increased risk of suicide, as even the referral for a diagnostic workup based on the results of the brief testing can be devastating to the patient. Preventive measures focusing on suicidal ideation must remain a focal point of patient care during

both the screening and diagnostic processes, especially when administering tests with increased levels of sensitivity.

Ultimately, these results suggest that there may be room for improvement of the utility of brief screening tests of cognition for dementia in older adults. Increasing the sensitivity of these brief cognitive screening measure for dementia would serve as a great contribution to the worldwide push for increased screening and assessment of cognitive abilities in older adults to allow for earlier detection of deficit associated with abnormal cognitive aging, such as MCI or AD. Increasing the ability of these measures to better detect cognitive deficit associated with abnormal aging in this population would indirectly result in increased referrals for full neuropsychological evaluations, which would subsequently result in the implementation of earlier treatment planning.

One cannot ignore the plethora of literature highlighting the effect of early intervention and treatment planning on increased quality of life in adults with cognitive impairment (Appels & Scherder, 2010; Giebel et al., 2014; Liu-Seifert et al., 2014). Should the results of the current study be replicated across different samples and clinical populations, this study may be among the first providing an alternative way of accomplishing that goal of expanding and improving upon currently existing early detection methods, which could potentially serve to be greatly beneficial to the healthcare industry.

Table 1

Distribution of DRS-2 Subscale Items and Scores

Subscale	# Questions	Relative %	# Points	Relative %
ATT	8	22	37	25
MEM	5	13	25	17
I/P	11	30	37	25
CONCEP	6	16	39	27
CONST	6	16	6	4
Total	36	100	144	100

Note. ATT = Attention. MEM = Memory. I/P = Initiation/Perseveration. CONCEP = Conceptualization. CONST = Construction.

Table 2

Original DRS-2 Subscale Weights

Subscale	# Qs	x	Logit Weight	=	# Total Pts
ATT	8	x	4.63	=	37
I/P	11	x	3.36	=	37
CONC	6	x	6.50	=	39
MEM	5	x	5.00	=	25
CONST	6	x	1.00	=	6
Total	36	–	–	–	144

Note. ATT = DRS-2 Attention subscale. I/P = DRS-2 Initiation/Perseveration subscale. CONCEPT = DRS-2 Conceptualization subscale. MEM = DRS-2 Memory subscale. CONST = DRS-2 Construction subscale.

Table 3

Means and Standard Deviations of Demographic Variables

	AD (n = 55)		MCI (n = 58)		Total (N = 113)	
Variable	M	SD	M	SD	M	SD
Age (yrs.)*	79.00	6.12	76.81	5.17	77.88	5.73
Education (yrs.)	14.95	3.10	14.91	3.29	14.93	3.19

Note: * Statistically significant group differences (p > .05, independent samples t-test).

Table 4

Means and Standard Deviations of Predictors as a Function of Group

	AD (n = 55)		MCI (n = 58)		Total (N = 113)	
Predictor	M	SD	M	SD	M	SD
ATT	31.84	3.65	33.76	2.75	32.82	3.35
I/P	29.44	6.11	34.36	3.237	31.96	5.43
CONC	33.24	4.12	35.62	2.72	34.46	3.66
MEM	15.05	3.03	20.76	3.03	17.98	4.16
CONST	5.58	0.81	5.83	0.60	5.71	0.72

Note. ATT = DRS-2 Attention subscale. I/P = DRS-2 Initiation/Perseveration subscale. CONC = DRS-2 Conceptualization subscale. MEM = DRS-2 Memory subscale. CONST = DRS-2 Construction subscale. AD = Alzheimer's disease. MCI = Mild Cognitive Impairment.

Table 5

Intercorrelations for Demographic and Predictor Variables

Variable	Age	Education	ATT	I/P	CONCEP	MEM	CONST
Age	1.0						
Education	.09	1.0					
ATT	-.02	-.04	1.0				
I/P	-.07	.20*	.25**	1.0			
CONC	-.09	.34*	.11	.22*	1.0		
MEM	-.13	.17	.16	.39**	.27**	1.0	
CONST	.02	.14	.13	.03	.15	.13	1.0

Note. ATT = DRS-2 Attention subscale. I/P = DRS-2 Initiation/Perseveration subscale. CONC = DRS-2 Conceptualization subscale. MEM = DRS-2 Memory subscale. CONST = DRS-2 Construction subscale.

* $p < .05$. ** $p \leq .01$.

Table 6

Full Model Logistic Regression Analysis Summary

Variable	*b*	SE	95% CI	Wald	df	p	OR	95% CI
ATT	.171	6.295	[-.019, .362]	3.096	1	.078	1.187	[.981, 1.436]
I/P	.186	.097	[.038, .336]	6.086	1	.014	1.204	[1.039, 1.396]
CONC	.157	.102	[-.042, .356]	2.406	1	.121	1.171	[.959, 1.428]
MEM	.495	.100	[.299, .692]	24.383	1	.000	1.641	[1.348, 1.998]
CONST	.408	.387	[-.351, 1.167]	1.110	1	.292	1.504	[.704, 3.213]
Constant	-28.192	6.295	–	20.058	1	–	–	–

Note. ATT = DRS-2 Attention subscale. I/P = DRS-2 Initiation/Perseveration subscale. CONC = DRS-2 Conceptualization subscale. MEM = DRS-2 Memory subscale. CONST = DRS-2 Construction subscale. Constant = Intercept.

Table 7

Application of Logit-Based DRS-2 Subscale Weighting Algorithm

Subscale	# Qs	x	Logit Weight	=	# Total Pts
ATT	8	x	.573	=	4.584
I/P	11	x	1.008	=	11.088
CONC	6	x	.576	=	3.456
MEM	5	x	2.062	=	10.31
CONST	6	x	0.292	=	1.752
Total	36	–	–	–	31.19

Note. ATT = DRS-2 Attention subscale. I/P = DRS-2 Initiation/Perseveration subscale. CONC = DRS-2 Conceptualization subscale. MEM = DRS-2 Memory subscale. CONST = DRS-2 Construction subscale.

Table 8
Calculation of Metric-Consistent Newly Computed DRS-2 Subscale Scores

Version	Subscale	# Pts	/	Tot Scr	=	Subscale %	x	Tot Scr	=	Weighted Tot
Original	CONC	39	/	144	=	.2708	–	–	–	–
	ATT	37	/	144	=	.2569	–	–	–	–
	I/P	37	/	144	=	.1736	–	–	–	–
	MEM	25	/	144	=	.2569	–	–	–	–
	CONST	6	/	144	=	.0416	–	–	–	–
	Total	144	–	–	–	–	–	–	–	–
Weighted	CONC	3.456	/	31.19	=	.1108	x	144	=	15.96
	ATT	4.584	/	31.19	=	.1470	x	144	=	21.16
	I/P	11.088	/	31.19	=	.3555	x	144	=	51.20
	MEM	10.31	/	31.19	=	.3306	x	144	=	47.60
	CONST	1.752	/	31.19	=	.0562	x	144	=	8.09
	Total	31.19	–	–	–	–	–	–	–	144

Note. ATT = DRS-2 Attention subscale. I/P = DRS-2 Initiation/Perseveration subscale. CONCEPT = DRS-2 Conceptualization subscale. MEM = DRS-2 Memory subscale. CONST = DRS-2 Construction subscale.

Outcome of the screening test	AD Diagnosis		Row Total
	Positive	Negative	
Positive	TP 43	FP 8	51
Negative	FN 12	TN 50	62
Column Total	55	58	113
Sensitivity = TP/(TP+FN) Specificity = TN/(TN+FP)			

Figure 1. *Summary of Original DRS-2 Sensitivity-Specificity Analysis*

Outcome of the screening test	AD Diagnosis		Row Total
	Positive	Negative	
Positive	(TP) 50	(FP) 15	65
Negative	(FN) 5	(TN) 43	48
Column Total	55	58	113
Sensitivity = TP/(TP+FN) Specificity = TN/(TN+FP)			

Figure 2. *Summary of Logit-Weighted DRS-2 Sensitivity-Specificity Analysis*

VII. References

Aerts, L., Heffernan, M., Kochan, N.A., Crawford, J.D., Draper, B., Trollor, J.N., Sachdev, P.S., & Brodaty, H. (2017). Effects of MCI subtype and reversion on progression to dementia in a community sample. *Neurology, 88*(23), 2225-2232.

Alzheimer's Association. (2017). 2017 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 13*(4), 325-373. https://doi.org/10.1016/j.jalz.2017.02.001

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington DC: American Psychiatric Association.

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd revised ed.). Washington DC: American Psychiatric Association.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington DC: American Psychiatric Association.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed.). Arlington VA: American Psychiatric Association.

Appels, B.A. & Scherder, E. (2010). The diagnostic accuracy of dementia-screening instruments with an administration time of 10 to 45 minutes for use in secondary care: A systematic review. *American Journal of Alzheimer's Disease & Other Dementias, 24*(4), 301-316.

Baldwin, P. (2015). Weighting components of a composite score using naïve expert judgments about their relative importance. *Applied Psychological Measurement, 39*(7), 539-550.

Baudic, S., Barba, G.D., Thibaudet, M.C., Smaghe, A., Remy, P., & Traykov, L. (2006). Executive function deficits in early Alzheimer's disease and their relations with episodic memory. *Archives of Clinical Neuropsychology, 21*(1), 15-21.

Blagosklonny, M.V. (2013). Aging is not programmed. *Cell Cycle, 12*(24), 3736-3742.

Bobholz, J.H., & Brandt, J. (1993). Assessment of cognitive impairment: Relationship of the Dementia Rating Scale to the Mini-Mental State Examination. *Journal of Geriatric Psychiatry and Neurology, 6*(4), 210-213.

Boller, F. & Forbes, M.M. (1998). History of dementia and dementia in history: An overview. *Journal of the Neurological Sciences, 158,* 125-133.

Borson, S., Frank, L., Bayley, P.J., Boustani, M., Dean, M., Lin, P.J., … & Ashford, J.W. (2013). Improving dementia care: The role of screening and detection of cognitive impairment. *Alzheimer's & Dementia, 9*(2), 151-159.

Boustani, M., Peterson, B., Hanson, L., Harris, R., & Lohr, K.N. (2003). Screening for dementia in primary care: A summary of evidence for the U.S. preventive services task force. *Annals of Internal Medicine, 138*, 927-937.

Bradford, A., Kunik, M.E., Schulz, P., Williams, S.P., & Singh, H. (2009). Missed and delayed diagnosis of dementia in primary care: Prevalence and contributing factors. *Alzheimer Disease and Associated Disorders, 23*(4), 306-314.

Bruscoli, M. & Lovestone, S. (2004). Is MCI really just early dementia? A systematic review of conversion studies. *International Psychogeriatrics, 16*(2), 129-140.

Burt, C. (1950). The influence of differential weighting. *British Journal of Statistical Psychology, 3*, 105-125.

Calderon, J., Perry, R.J., Erzinclioglu, S.W., Berrios, G.E., Dening, T.R., & Hodges, J.R. (2001). Perception, attention, and working memory are disproportionately impaired in dementia with Lewy bodies compared with Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *70*(2), 157-164.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology, 21*(3), 217-227.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Acadmic.

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* (3rd Eds.). Hillsdale, Erlbaum.

Colantonio, A., Becker, J.T., & Huff, F.J. (1993). Factor structure of the Mattis Dementia Rating Scale among patients with probable Alzheimer's disease. *The Clinical Neuropsychologist, 7*(3), 313-318.

Collette, F., Amieva, H., & Adam, S., Hogge, M., Van der Linden, M., Fabrigoule, C., & Salmon, E. (2007). Comparison of inhibitory functioning in mild Alzheimer's disease and frontotemporal dementia. *Cortex, 43,* 866-874.

Crowell, T.A., Luis, C.A., Vanderploeg, R.D., Schinka, J.A., & Mullan, M. (2002). Memory patterns and executive functioning in mild cognitive impairment and Alzheimer's disease. *Aging Neuropsychology of Cognition, 9,* 288-297.

Danysz, W., Parsons, C.G., Mobius, H., Stoffler, A., & Quack, G. (2000). Neuroprotective and symptomatological action of memantine relevant for

Alzheimer's disease – A unified glutamatergic hypothesis on the mechanism of action. *Neurotoxicity Research, 2*(2-3), 85-97.

Draper, B., Peisah, C., Snowdon, J., & Brodaty, H. (2010). Early dementia diagnosis and the risk of suicide and euthanasia. *Alzheimer's & Dementia, 6*(1), 75-82.

Emrani, S., Libon, D.J., Lamar, M., Price, C.C., Jefferson, A.L., Gifford, K.A., … & Au, R. (2018). Assessing working memory in mild cognitive impairment with serial order recall. *Journal of Alzheimer's Disease, 61*(3), 917-928.

Erlangsen, A., Zarit, S.H., & Conwell, Y. (2008). Hospital-diagnosed dementia and suicide: A longitudinal study using prospective, nationwide register data. *The American Journal of Geriatric Psychiatry, 16*(3), 220-228.

Fengler, S., Kessler, J., Timmermann, L., Zapf, A., Elben, S., Wojtecki, L… & Kalbe, E. (2016). Screening for cognitive impairment in Parkinson's disease: Improving the diagnostic utility of the MoCA through subtest weighting. *PLoS ONE, 11*(7).

Fischer, P., Jungwirth, S., Zehetmayer, S., Weissgram, S., Hoenigschnabl, S., Gelpi, E., … Tragl, K.H. (2007). Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology, 68,* 288-291.

Folstein, M.F., Folstein, S.E., & McHugh, P.R.. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

Giebel, C.M., Sutcliffe, C., Stolt, M., Karlsson, S., Renom-Guiteras, A., Soto, M., Verbeek, H… & Challis, D. (2014). Deterioration of basic activities of daily living and their impact on quality of life across different cognitive stages of dementia: A European study. *International Psychogeriatrics, 26*(8), 1283-1293.

Goldsmith, T.C. (2012). On the programmed/non-programmed aging controversy. *Biochemistry, 77*(7), 729-732.

Govindarajulu, Z. (1988). Alternative methods for combining several test scores. *Educational and Psychological Measurement, 48,* 53-60.

Hashimoto, M., Kazui, H., Matsumoto, K., Nakano, Y., Yasuda, M., & Mori, E. (2005). Does donepezil treatment slow the progression of hippocampal atrophy in patients with Alzheimer's disease? *American Journal of Psychiatry, 162*(4), 676-682.

Heller, T., Scott, H.M., Janicki, M.P., & Pre-Summit Workgroup on Caregiving and Intellectual and Developmental Disabilities. (2018). Caregiving, intellectual disability, and dementia: Report of the Workgroup on Caregiving and Intellectual and Developmental Disabilities. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 1-11.

IBM Corp. Released 2017. IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY: IBM Corp.

Iliffe, S., Manthorpe, J., & Eden, A. (2003). Sooner or later? Issues in the early diagnosis of dementia in general practice: A qualitative study. *Family Practice, 20*(4), 376-381.

Imtiaz, B., Tolppanen, A., Kivipelto, Miia, & Soininen, K. (2014). Future directions in Alzheimer's disease from risk factors to prevention. *Biochemical Pharmacology, 88*(4), 661-670.

Ivnik, R.J., Malec, J.F., Tangalos, E.G., Petersen, R.C., Kokmen, E., & Kurland, L.T. (1990). The Auditory-Verbal Learning Test (AVLT): Norms for ages 55 years and older. *A Journal of Counseling and Clinical Psychology, 2*(3), 304-312.

Ivnik, R.J., Smith, G.E., Tangalos, E.G., Petersen, R.C., Kokmen, E., & Kurland, L.T. (1991). Wechsler Memory Scale: IQ-dependent norms for persons ages 65 to 97 years. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*(2), 156-161.

Ivnik, R.J., Malec, J.F., Smith, G.E., Tangalos, E.G., Petersen, R.C., Kokmen, E., & Kurland, L.T. (1992). Mayo's older americans normative studies: WAIS-R norms for ages 56 to 97. *The Clinical Neuropsychologist, 6*(S1), 1-30.

Jack, C.R., Jagust, W.J., Shaw, L.M., Aisen, P.A., Weiner, M.W., Petersen, R.C., & Trojanowski, J.Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurology, 9,* 119-128.

Jekel, K., Damian, M., Wattmo,, C., Hausner, L., Bullock, R., Connelly, P.J., Dubois, B… & Frolich, L. Mild cognitive impairment and deficits in instrumental activities of daily living: A systematic review. *Alzheimer's Research & Therapy, 7*(17).

Jin, K. (2010). Modern biological theories of aging. *Aging and Disease, 1*(2), 72-74.

John, S.E., Grunani, A.S., Bussell, C., Saurman, J.L., Griffin, J.W., & Gavett, B.E. (2016). The effectiveness and unique contribution of neuropsychological tests and the δ latent phenotype in the differential diagnosis of dementia in the uniform data set. *Neuropsychology, 30*(8), 946-960.

Jurica, P.J., Leittenn, C.L., & Mattis, S. (2001). *Dementia Rating Scale-2: Professional Manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Karantzoulis, S. & Galvin, J.E. (2011). Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Review of Neurotherapeutics, 11*(11), 1579-1591.

Kessler, H.R., Roth, D.L., Kaplan, R.F., & Goode, K.T. (1994). Confirmatory factor analysis of the mattis dementia rating scale. *The Clinical Neuropsychologist, 8*(4), 451-461.

Kirkwood, T.B.L. & Melov, S. (2011). On the programmed/non-programmed nature of ageing within the life history. *Current Biology, 21*, R701-R707.

Kochanek, K.D., Murphy, S.L., Xu, J.Q., & Tejada-Vera, B. Deaths: Final data for 2014. (2016). *National Vital Statistics Reports, 65*(4), 1-122.

Kowald, A. & Kirkwood, T.B.L. (2016). Can aging be programmed? A critical literature review. *Aging Cell, 15*, 986-998.

Krishnan, K.R., Charles, H.C., Doraiswamy, P.M., Mintzer, J., Weisler, R., Yu, X., Perdomo, C., Ieni, J.R., & Rogers, S. (2003). Randomized, placebo-controlled trial of the effets of donepezil on neuronal markers and hippocampal volumes in Alzheimer's disease. *American Journal of Psychiatry, 160*(11), 2003-2011.

Kubu, C.S., Ready, R.E., Festa, J.R., Roper, B.L., & Pliskin, N.H. (2016). The times they are a changin': Neuropsychology and integrated care teams. *The Clinical Neuropsychologist, 30*(1), 51-65. DOI: 10.1080/13854046.2015.1134

Lezak, M., Howieson, D., Bigler, E., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York: Oxford University Press.

Li, J.Q., Tan, L., Wang, H.T., Tan, M.S., Tan, L., Xi, W., … & Yu, J.T. (2016). Risk factors for predicting progression from mild cognitive impairment to Alzheimer's

disease: A systematic review and meta-analysis of cohort studies. *Journal of Neurology, Neurosurgery, and Psychiatry, 87,* 476-484.

Libon, D.J., McMillan, C.D., Gunawardena, C., Powers, L., Massimo, A., Khan, B., … & Grossman, M. (2009). Neurocognitive contributions to verbal fluency deficits in frontotemporal lobar degeneration. *Neurology, 73*(3), 535-542.

Liu-Seifert, H., Siemers, E., Sundell, K., Price, K., Han, B., Selzler, K., Aisen, P… & Mohs, R. (2015). Cognitive and functional decline and their relationship in patients with mild Alzheimer's dementia. *Journal of Alzheimer's Disease, 43*(3), 949-955.

Loring, D.W. (2015). *INS dictionary of neuropsychology & clinical neurosciences.* Oxford, UK: Oxford University Press.

Lucas, J.A., Ivnik, R.J., Smith, G.E., Bohac, D.L., Tangalos, E.G., Kokmen, E., Graff-Radford, N.R., & Petersen, R.C. (1998). Normative data for the Mattis Dementia Rating Scale. *Journal of Clinical and Experimental Neuropsychology, 20*(4), 536-547.

Lucas, J.A., Ivnik, R.J., Willis, F.B., Ferman, T.J., Smith, G.E., Parfitt, F.C., Petersen, R.C., & Graff-Radford, N.R. (2005). MAYO's older African American normative studies: Normative data for commonly used clinical neuropsychological measures. *The Clinical Neuropsychologist, 19,* 162-183.

Marcopulos, B.A., McLain, C.A., & Giuliano, A.J. (1997). Cognitive impairment or inadequate norms? A study of healthy, rural, older adults with limited education. *The Clinical Neuropsychologist, 11*(2), 111-131.

Marshall, G.A., Rentz, D.M., Frey, M.T., Locascio, J.J., Johnson, K.A., & Sperling, R.A. (2012). Executive function and instrumental activities of daily living in MCI and AD. *Alzheimers Dementia, 7*(3), 300-308.

Martyr, A. & Clare, L. (2012). Executive function and activities of daily living in Alzheimer's disease: A correlational meta-analysis. *Dementia and Geriatric Cognitive Disorders, 33,* 189-203.

Mattis, S. (1976). Mental status examination for organic mental syndrome in the elderly patient. In: L. Bellak & T.B. Karasu (Eds.), *Geriatric psychiatry* (pp. 77-121). New York: Grune & Stratton.

Mattis, S. (1988). *Dementia Rating Scale. Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Maynard, S., Fang, E.F., Scheibye-Knudsen, M., Croteau, D.L., & Bohr, V.A. (2015). DNA damage, DNA repair, aging, and neurodegeneration. *Cold Spring Harb Perspect Med, 5*, a025130.

McKhann, G.M., 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*(7), 939-944.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., 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 & Dementia, 7*, 263-269.

Meehl, P.E. & Rose, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*(3), 194-216).

Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., Dies, R.R., Eisman, E.J., Kubiszyn, T.W., & Reed, G.M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*(2), 128-165.

Miller, L.J., Myers, A., Prinzi, L., & Mittenberg, W. (2009). Changes in intellectual functioning associated with normal aging. *Archives of Clinical Neuropsychology, 24*, 681-688.

Mitchell, A.J. & Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robus inception cohort studies. *Acta Psychiatrica Scandinavica, 119*(4), 252-265.

Mitrushina, M. (2009). Cognitive screening methods. In: Grant, I. & Adams, K.M. (Eds.), *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders* (pp. 101-126). New York, New York: Oxford University Press.

Monsch, A.U., Bondi, M.W., & Salmon, D.P. (1995). Clinical validity of the Mattis Dementia Rating Scale in detecting dementia of the Alzheimer type: A double cross-validation and application to a community-dwelling sample. *Archives of Neurology, 52*(9), 899-904.

Montgomery, K.M. & Costa, L. (1983). Neuropsychological test performance of a normal elderly sample. Paper presented at the annual meeting of the International Neuropsychological Society, Mexico City, Mexico.

Nadler, J.D., Mittenberg, W., DePiano, F.A., & Schneider, B.A. Effect of patient age on neuropsychological test interpretation. *Professional Psychology: Research and Practice, 25*(3), 288-295.

Nestor, P.J., Scheltens, P., & Hodges, J.R. (2004). Advances in the early detection of Alzheimer's disease. *Nature Medicine, 5*, s34-41.

Owsley, C., Sloane, M., McGwin, G., & Ball, K. (2002). Timed instrumental activities of daily living tasks: Relationship to cognitive function and everyday performance assessments in older adults. *Gerontology*, *48*(4), 254-265.

Palmer, K., Backman, L., Winblad, B., & Fratiglioni, L. (2003). Detection of Alzheimer's disease and dementia in the preclinical phase: Population based cohort study. *BMJ: British Medical Journal*, *326*(7383), 245.

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R., & Feinstein, A.R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology, 49*(12), 1373-1379.

Perloff, J.M. & Persons, J.B. (1988). Biases resulting from the use of indexes: An application to attributional style and depression. *Psychological Bulletin, 103*, 95-104.

Peter, J.P. (1979). Reliability: A review of psychometric basics and recent marketing practices. *American Marketing Association, 16*(1), 6-17.

Petersen, R.C., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Schaid, D.J., Thibodeau, S.N., … & Kurland, L.T. (1995). Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals. *JAMA, 273*(16), 1274-1278.

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, 53,* 303-308.

Petersen, R.C., Stevens, J.C., Ganguli, M., Tangalos, E.G., Cummings, J.L., & DeKosky, S.T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review). *Neurology, 56,* 1133-1142.

Petersen, R.C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: A concept in evolution. *Journal of Internal Medicine, 275,* 214-228.

Possin, K.L. (2010). Visual spatial cognition in neurodegenerative disease. *Neurocase, 16*(6), 466-487.

Ramsay, M.C. & Reynolds, C.R. (2000). Development of a scientific test: A practical guide. In: Goldstein, G. & Hersen, M. (3rd Eds.), *Handbook of Neuropsychological Assessment.* (pp. 21-42). Kidlington, Oxford: Elsevier Science Ltd.

Rascovsky, K. & Grossman, M. (2013). Clinical diagnostic criteria and classification controversies in frontotemporal lobar degeneration. *International Review of Psychiatry, 25*(2), 145-158.

Retzlaff, P.D. & Gibertini, M. (2014). Neuropsychometric issues and problems. In: Vanderploeg, R.D. (Ed.), *Clinician's Guide to Neuropsychological Assessment*.

Risacher, S.L., Saykin, A.J., West, J.D., Shen, L., Firpi, H.A., McDonald, B.C., & the ADNI. (2009). Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Current Alzheimer Research, 6,* 347-361.

Ritter, A.R., Leger, G.C., Miller, J.B., & Banks, S.J. Neuropsychological testing in pathologically verified Alzheimer disease and Frontotemporal dementia: How well do the uniform data set measures differentiate between diseases? *Alzheimer Disease & Associated Disorders, 31*(3), 187-191.

Rizzo, M., Anderson, S.W., Dawson, J., & Nawrot, M. Vision and cognition in Alzheimer's disease. *Neuropsychologia, 38*(8), 1157-1169.

Roundtree, S.D., Chan, W., Pavlik, V.N., Darby, E.J., Siddiqui, S., & Doody, R.S. (2009). Persistent treatment with cholinesterase inhibitors and/or memantine slows clinical progression of Alzheimer's disease. *Alzheimer's Research & Therapy, 1*(7).

Royall, D.R., Palmer, R., Chiodo, L.K., & Polk, M.J. (2004). *Journal of the American Geriatrics Society, 52*(3), 346-352.

Sacco, G., Joumier, V., Darmon, N., Dechamps, A., Derreumaux, A., Lee, J.H., Piano, J… & Robert, P. (2012). Detection of activities of daily living impairment in Alzheimer's disease and mild cognitive impairment using information and communication technology. *Clinical Interventions in Aging, 7*, 539-549.

Salmon, D.P., Thal, L.J., Butters, N., & Heindel, W.C. (1990). Longitudinal evaluation of dementia of the Alzheimer type: A comparison of 3 standardized mental status examinations. *Neurology, 40*(8), 1225-1230.

Scheiber, C., Chen, H., Kaufman, A.S., & Weiss, L.G. (2017). How much does WAIS-IV perceptual reasoning decline across the 20 to 90-year lifespan when processing speed is controlled? *Applied Neuropsychology: Adult, 24*(2), 116-131.

Schoenberg, M.R. & Scott, J.G. (2011). The neuropsychology referral and answering the referral question. In: Schoenberg, M.R. & Scott, J.G. (Eds.), *The little black book of neuropsychology* (pp. 1-37). New York, NY: Springer.

Schoenberg, M.R., & Duff, K. (2011). Dementias and mild cognitive impairment in adults. In: Schoenberg, M.R. & Scott, J.G. (Eds.), *The little black book of neuropsychology* (pp. 1-37). New York, NY: Springer.

Shany-Ur, T., Lin, N., Rosen, H.J., Sollberger, M., Miller, B.L., & Rankin, K.P. (2014). Self-awareness in neurodegenerative disease relies on neural structures mediating reward-driven attention. *Brain: A Journal of Neurology, 137*, 2368-2381.

Singer, R. (2011). Neurotoxicity in neuropsychology. In: Schoenberg, M. R. & Scott, J. G., (Eds.), *The little black book of neuropsychology* (pp. 813-838).

Smith, G.E. & Bondi, M.W. (2008). Normal aging, mild cognitive impairment and Alzheimer's disease. In: Morgan, J.E. & Ricker, J.H. (Eds.), *Textbook of clinical neuropsychology*, (pp. 768-780). New York: Psychology Press.

Smith, M.J., Breitbart, W.S., & Platt, M.M. (1995). A critique of instruments and methods to detect, diagnose, and rate delirium. *Journal of Pain and Symptom Management, 10*(1), 35-77.

Stern, Y., Andrews, H., Pittman, J., Sano, M., Tatemichi, T., Lantigua, R., & Mayeux, R. (1992). Diagnosis of dementia in a heterogeneous population. *Archives of Neurology, 49*, 453-460.

Tabachnick, B.G. & Fidell, L.S. (2013). *Using Multivariate Statistics* (6th edition). Boston, MA: Pearson.

Tabuas-Pereira, M., Baldeiras, I., Duro, D., Santiago, B., Ribeiro, M.H., Leitao, M.J., Oliveira, C., & Santana, I. (2016). Prognosis of early-onset vs. late-onset mild cognitive impairment: Comparison of conversion rates and its predictors. *Geriatrics, 1*(11), 1-12.

Tartaglia, M.C., Rosen, H.J., & Miller, B.L. (2011). Neuroimaging in dementia. *Neurotherapeutics: The Journal of the American Society for Experimental Neurotherapeutics, 8*, 82-92.

Thompson, G.H. (1940). Weighting for battery reliability and prediction. *British Journal of Psychology, 30*, 357-366.

Tombaugh, T.N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14*(2), 167-177).

Vangel, S.J. & Lichtenberg, P.A. (1995). Mattis dementia rating scale: Clinical utility and relationship with demographic variables. *The Clinical Neuropsychologist, 9*(3), 209-213.

Vanderploeg, R.D. (2014). *Clinician's Guide to Neuropsychological Assessment.* (2nd Eds.). New York, New York: Routledge.

Vincent, G.K., & Velkof, V.A. (2010). The next four decades: The older population in the United States: 2010 to 2050. Washington, DC: U.S. Census Bureau.

Wang, M., & Stanley, J. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research, 40*, 663-705.

Wilkinson, D. (2012). A review of the effects of memantine on clinical progression in Alzheimer's disease. *International Journal of Geriatric Psychiatry, 5*(6).

Williams, D.R. (2006). Tauopathies: Classification and clinical update on neurodegenerative diseases associated with microtubule-associated protein tau. *Internal Medicine Journal, 36*(10), 652-660.

Wilson, R., Boyle, P., Yu, L., Barnes, L., Sytsma, J., Buchman, A., Bennett, D., & Schneider, J. (2015). Temporal course and pathologic basis of unawareness of memory loss in dementia. *Neurology, 85*(11), 984-991.

Wisdom, N.M., Mignogna, J., & Collins, R.L. (2012). Variability in Wechsler Adult Intelligence Scale-IV subtest performance across age. *Archives of Clinical Neuropsychology, 27*, 389-397.

Woodard, J.L., Salthouse, T.A., Godsall, R.E., & Green, R.C. (1996). Confirmatory factor analysis of the Mattis Dementia Rating Scale in patients with Alzheimer's disease. *Psychological Assessment, 8*(1), 85-91.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* Geneva: World Health Organization.