Psychometric Properties of the Saint Louis University Mental Status Examination (SLUMS) for the Identification of Mild Cognitive Impairment (MCI) in a Veteran Sample

Susan Stern

Follow this and additional works at: https://scholarworks.gsu.edu/psych_diss

Recommended Citation
https://scholarworks.gsu.edu/psych_diss/125

This Dissertation is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
The Saint Louis University Mental Status (SLUMS) Examination is a relatively new brief cognitive screening measure developed for use with veterans. To date, there has been a paucity of research on its psychometric properties. Using a sample of 148 male veterans referred to a VA Mild Cognitive Impairment (MCI) Clinic for evaluation, the SLUMS’ ability to discriminate between MCI versus other diagnoses or no diagnosis was compared to results from a more comprehensive neuropsychological battery. Approximately 51% of the sample was diagnosed with MCI, 16% with Major Depressive Disorder (MDD), 17% did not meet criteria for a diagnosis, and 16% were given some other DSM-IV-TR diagnosis.

The SLUMS demonstrated poor internal consistency (Cronbach’s alpha = .57), but scores were significantly correlated with scores on every neuropsychological measure, except for Trails B. Diagnostic discriminability was comparable to that of the more time intensive neuropsychological battery for discrim-
inating between MCI and no diagnosis, and MCI and MDD. In the current sample, a cutoff score of 25 was optimal for discriminating between MCI and no diagnosis, whereas a slightly lower cutoff score of 24 is recommended for discriminating between MCI and those with MDD. Diagnostic indicators were poor for the SLUMS and the battery when discriminating between MCI and a heterogeneous group of other disorders. Possible reasons for low reliability in such a screening measure in the context of convergent validity are discussed. It is concluded that the SLUMS may be a viable brief cognitive screening measure in such veteran populations, particularly when discriminating between MCI and MDD; however, additional studies should be completed to evaluate other forms of consistency, such as test-retest reliability.

INDEX WORDS: Cognitive screening, Mild cognitive impairment, SLUMS
PSYCHOMETRIC PROPERTIES OF THE SAINT LOUIS UNIVERSITY
MENTAL STATUS EXAMINATION (SLUMS) FOR THE IDENTIFICATION OF MILD COGNITIVE IMPAIRMENT (MCI) IN A VETERAN SAMPLE

by

Susan K. Stern

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2014
PSYCHOMETRIC PROPERTIES OF THE SAINT LOUIS UNIVERSITY
MENTAL STATUS EXAMINATION (SLUMS) FOR THE IDENTIFICATION OF MILD COGNITIVE
IMPAIRMENT (MCI) IN A VETERAN SAMPLE

by

Susan K. Stern, M.A.

Committee Chair: Robin D. Morris

Committee: Suzanne Penna
Erin Tone
Lindsey Cohen
David Washburn

Electronic Version Approved:
Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2014
DEDICATION

I would not be where I am today if it were not for my wonderfully supportive parents, Dr. Sam and Renee Stern, who provided me with everything I could have ever needed to succeed. I am forever grateful to them for their love, support and guidance. I would also like to thank my brothers, Steve and Matt, who have offered encouraging words to me whenever they could during my hardest times. I would also like to thank Jason Shwartz, who believed in me and my abilities and provided never-ending love, laughter, support, advice, and encouragement.

I would also like to thank Dr. Robin Morris for challenging me while always remaining supportive. He has been a wonderful advisor for the past three years. When we first agreed to work together, I shared my careers goals with him. He readily acknowledged and accepted my path and committed to supporting me in my success. He helped guide me in the right direction and make appropriate decisions regarding my research. I appreciated that there was rarely a meeting in which we did not share a good laugh!

This dissertation would not have been possible without Dr. Suzanne Penna. Suzanne was a fantastic supervisor, wonderful mentor, supportive committee member, and a joy to spend time with during a lunch break from work. She talked to me about work-life balance, particularly as a woman, which is a topic that is rarely discussed. Her zeal for humor and happiness was contagious, and I try to carry it with me as I continue on in my training.

Finally, I would like to thank my committee members, Drs. Suzanne Penna, Erin Tone, Lindsey Cohen, and David Washburn, for taking time from their very busy schedules to read my dissertation, offer their advice, wisdom, statistical and grammatical prowess, and attend my proposal and defense. Their thought-provoking and challenging questions have and will continue to help me develop as a scientist-practitioner.
ACKNOWLEDGEMENTS

I would like to acknowledge the support Atlanta Veterans Affairs Medical Center (VAMC). The Atlanta VAMC provided me with access to their clinical database, making this research possible.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. v

LIST OF TABLES .......................................................................................................................... viii

1 INTRODUCTION .......................................................................................................................... 1

1.1 What is Mild Cognitive Impairment? History, Definition, and Relevant Information ............ 2

1.2 Screening for MCI .................................................................................................................... 7

1.3 Veterans: A Unique Population ............................................................................................. 15

1.4 The Saint Louis University Mental Status Examination .................................................... 16

1.5 Proposed Study ....................................................................................................................... 25

2 METHODS .................................................................................................................................. 27

2.1 Procedures ............................................................................................................................... 27

2.2 Participants .............................................................................................................................. 30

2.3 Measures ................................................................................................................................ 32

2.3.1 Cognitive screeners ............................................................................................................ 33

2.3.2 Estimated intelligence ........................................................................................................ 33

2.3.3 Attention ............................................................................................................................. 33

2.3.4 Processing speed ............................................................................................................... 34

2.3.5 Language ............................................................................................................................ 34

2.3.6 Memory ............................................................................................................................... 35

2.3.7 Executive functioning ....................................................................................................... 36

2.3.8 Emotional functioning ...................................................................................................... 36

2.4 Confirmatory Factor Analysis of the Neuropsychological Battery ....................................... 36
2.5 Classification of Subjects .................................................................37
2.6 Analyses ..........................................................................................40

3 RESULTS ..........................................................................................42
3.1 Data Screening ................................................................................42
3.2 Reliability .......................................................................................43
3.3 ROC Curve Analysis for the SLUMS ...............................................44
  3.3.1 MCI vs. No Diagnosis .................................................................44
  3.3.2 MCI vs. MDD ...........................................................................44
  3.3.3 MCI vs. Other Diagnosis ..........................................................44
3.4 ROC Curve Analysis for the Neuropsychological Battery ..................45
  3.4.1 MCI vs. No Diagnosis .................................................................45
  3.4.2 MCI vs. MDD ...........................................................................45
  3.4.3 MCI vs. Other Diagnosis ..........................................................45
3.5 ROC Curve Analysis for the Factor Scores ........................................46

4 DISCUSSION .....................................................................................47
4.1 Clinical Implications ........................................................................54

5 REFERENCES ....................................................................................58

6 APPENDIX A ....................................................................................75

7 APPENDIX B ....................................................................................76
LIST OF TABLES

Table 1 DSM-IV-TR and NINCDS-ADRDA Criteria for Diagnosing Dementia and Alzheimer's Disease ................................................................. 79
Table 2 Sample Demographics ......................................................................................................................... 80
Table 3 Correlations Between SLUMS Score and Demographic Factors and Neuropsychological Test Scores ................................................................................................. 81
Table 4 Inter-Item Correlations for the SLUMS ................................................................................................. 83
Table 5 Item-Total Statistics for the SLUMS ................................................................................................. 84
Table 6 Mean Sample Score on Each Neuropsychological Measure ................................................................. 85
Table 7 Cutoff Scores for Mild Cognitive Impairment on the SLUMS ................................................................. 86
Table 8 Cutoff Scores (Number of Deficits) for MCI using a Neuropsychological Battery ......................... 87
Table 9 Cutoff Scores (Number of Deficits for MCI Using Factor Scores ......................................................... 88
Table 10 Diagnostic Indicators for the SLUMS, the Comprehensive Neuropsychological Battery, the Factor Scores, and the DRS-2 ........................................................................... 89
LIST OF FIGURES

Figure 1 Initial Confirmatory Factor Analysis Model ................................................................. 90

Figure 2 Final Confirmatory Factor Analysis Model ................................................................. 91
1 INTRODUCTION

As our population ages, the risk of developing cognitive impairment and dementia increases. The risk of cognitive impairment and dementia at older ages is especially relevant to the veteran population, as this group is older than the non-veteran population (median age for veterans in 2009 was 64 years old, while the median age for non-veterans was 49 years old; Agha, Lofgren, VanRuiswyk, & Layde, 2000; National Center for Veterans Analysis and Statistics, 2011b). Research and clinical interventions have begun to focus on trying to identify the earliest stages of neurocognitive disease. In such patients, cognitive changes are observed, but daily functioning is not significantly impacted. This condition has become known as mild cognitive impairment (MCI). Early identification of such subtle changes are thought to provide the best opportunity to intervene early in the disease process and potentially delay or reverse a neurodegenerative process, particularly if treatable etiologies are identified.

Characterizing the signs of such early deficits may also help determine a patient’s ultimate diagnosis or prognosis (Petersen, 2004; Ravaglia et al., 2006). The ability to assess such deficits reliably could help identify those early treatments or preventative measures in dementing processes. There is some early evidence that certain medications can provide small improvements in cognition and activities of daily living (ADLs) or delay the progression of dementia that is detected early (Birks, 2006; Petersen, Thomas, Grundman, Bennet, Doody, & Ferri, 2005; Van Dam & De Deyn, 2006). There is also the possibility that some risk factors can also be modified to potentially reverse early symptoms (e.g., Etgen, Sander, Bickel, & Först, 2011). However, there is still significant difficulty in the ability to identify accurately the prodromal stages of a dementing process.

A comprehensive neuropsychological assessment is a well-validated tool to identify accurately a dementing process, even more so if serial assessments are completed. Such assessments can also be used to rule out other possible explanations. However, such comprehensive assessments are both very
time consuming and costly. Particularly when trying to identify early stage cognitive pathology in an increasingly aging society, the potential cost in time and money being used on such thorough evaluations of clinically “normal” or non-dementing persons is high, especially given the prevalence of such disorders in the general population. Because of this, more time-efficient, sensitive, and specific screening measures are needed.

A new neurocognitive screener, the Saint Louis University Mental Status Examination (SLUMS; Tariq, Tumosa, Chibnall, Perry, & Morely, 2006) purports to measure sensitively such early changes in veterans, but it has not been thoroughly validated with regard to its ability to detect MCI. This study adds to the knowledge about this screening measure by providing a study of its psychometric and screening capabilities in an independent sample of veterans referred for cognitive evaluation.

1.1 What is Mild Cognitive Impairment? History, Definition, and Relevant Information

The nomenclature for the transitional status between normal cognition and dementia has evolved over the years. Kral (1962) first described benign and malignant senescent forgetfulness; the former referred to the occasional difficulty older adults experience in recalling unimportant data from an event, whereas the latter referred to a more persistent deficit in encoding, storage, or retrieval of unimportant and important information from the recent past. The malignant form is progressive and associated with a significantly higher mortality rate than the benign form.

Since then, various nomenclatures have included terms such as “mild cognitive impairment,” “dementia prodrome,” “incipient dementia,” “isolated memory impairment,” “preclinical Alzheimer’s disease,” and “cognitive impairment not dementia” (Allegri, Glaser, Taragano, & Buschke, 2008; Petersen, 2004; Sperling et al., 2011). The most commonly accepted term, and the term that will be used in this study, is mild cognitive impairment (MCI). Petersen and colleagues described the clinical characterization of MCI and determined criteria for its diagnosis (Petersen, Smith, Waring, Ivnik, Tangalos, & Kokmen, 1999). In their initial characterization of the disorder, they claimed that it was a prodromal
state that developed into Alzheimer’s disease (AD). Thus, the initial criteria only included a subjective complaint and objective deficits in memory, with no consideration for deterioration in other cognitive domains. The criteria required (1) the presence of a memory complaint, (2) normal activities of daily living, (3) normal general cognitive functioning, (4) abnormal memory for age, and (5) they are not demented.

Since that time, however, it has been suggested that there may be several subtypes of MCI. Longitudinal studies have demonstrated that suggested subtypes may be associated with different etiologies and may evolve into various diseases. For example, single-domain amnestic MCI (a-sd-MCI; a deficit in only memory) may reflect prodromal AD or depression, whereas single-domain non-amnestic MCI (na-sd-MCI; i.e., only visuospatial, executive functioning, or language, etc.) may reflect the early stage of dementia with Lewy bodies (DLB), frontotemporal dementia, vascular dementia (VaD), or primary progressive aphasia. Similarly, a multi-domain amnestic MCI (a-md-MCI) may reflect the impact of depression or the early stages of AD or VaD, whereas a multi-domain non-amnestic MCI (na-md-MCI) is more likely to progress to DLB or VaD (Petersen, Doody, et al., 2001; Petersen, 2004).

Not all cases of MCI progress. Some individuals remain stable over time, and others recover. This may reflect variations in rate of progression, etiology, or possibly even diagnostic measurement error. Given the probability of developing AD and the high rates of vascular disease that tend to increase with age, poverty, and lower education in the United States (Pleis, Ward, & Lucas, 2010), it is likely that some profiles may reflect a mixed etiology.

To reflect the growing evidence of heterogeneity in the diagnosis of MCI, an expert symposium in 2004 led to the proposal of new criteria that allowed for the inclusion of impaired cognitive domains other than memory and allowed for mild deficits in daily functioning. The new criteria required that an individual (1) is not normal, but not demented (i.e., does not meet criteria for a dementing disorder according to Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition, Text Revision [DSM-IV-
TR; American Psychiatric Association, 2000] or ICD-10), (2) has preserved basic activities of daily living (ADLs)/minimal impairment in complex instrumental functions, and (3) there is a self- or informant-report of cognitive decline in conjunction with an impairment on objective cognitive measures and/or evidence of an objective decline over time on cognitive tasks (Winblad et al., 2004). These criteria did not specify time periods for observations of declines, what measures to use, or what constitutes an impairment, although 1 to 1.5 standard deviations (SD) below the normative mean is the most common (Budson & Solomon, 2012), and 1.5 SD has been the most commonly used cutoff score in research (Artero, Petersen, Touchon, & Ritchie, 2006). However, the designation of an “impairment” in a clinical setting is typically based on clinical judgment.

Utilizing the revised MCI criteria, the prevalence and rates of progression to dementia vary dramatically. A recent review (Petersen et al., 2009) presented prevalence rates from a number of studies worldwide, ranging from 3-24%. Artero and colleagues (2006) reported a prevalence rate of 17% in a French population-based sample. Incident rates also vary widely. In a systematic review, Luck, Luppa, Briel, & Riedel-Heller (2010) reported incidence rates of a-MCI combined (single and multi-domain) as ranging from 9.9 to 40.6 per 1,000 person-years. Incidence rates of na-MCI combined ranged from 28 to 36.3 per 1,000 person-years. Incidence of all MCI types ranged from 51 to 76.8 per 1,000 person-years.

Reports of rates of progression to dementia are more limited. For example, Petersen et al. (1999) at the Mayo Clinic reported that the annual rate of progression to dementia was approximately 12%, with 80% of the MCI sample converting to dementia over a six-year period (Petersen, Doody, et al., 2001). In comparison, the conversion rate in the general population is 1% to 2% per year. Additionally, Farias, Mungas, Reed, Harvey, and DeCarli (2009) reported that subjects from a clinic sample (memory clinic) progressed from MCI at baseline to dementia at a rate of 13% per year, whereas subjects from the community converted at a rate of 3% per year. Bruscoli and Lovestone (2004) conducted a meta-analysis of studies published between 1991 and 2001 and reported an overall conversion rate of 10%,
with self-referred clinic samples having the highest rates of conversion. These data suggest that having subjective or objective cognitive impairments severe enough to warrant professional evaluation increase the rate of decline to dementia.

Not all individuals with MCI progress. For example, Ravaglia and colleagues (2006) found that 29% of their sample progressed to dementia (71% AD, 29% VaD), whereas 67% remained stable, and 4% reverted back to normal. The authors explain the occurrence of stable or reversed cognitive impairments as possibly due to premorbid, long-standing poor cognitive functioning or the relatively short follow-up period (average of 2.8 years, SD=1.6, range 6 months to 5 years).

A number of factors are predictive of progression, including initial severity of cognitive impairment, apolipoprotein E4 (ApoE e4) carrier status, brain atrophy on MRI, cerebrospinal fluid (CSF) biomarkers compatible with AD, and more (Petersen et al., 1999). Interestingly, Gomar, Bobes-Bascaran, Conejero-Goldberg, Davies, and Goldberg (2011) investigated the predictive utility of combinations of biomarkers, cognitive markers, and functional assessment measures to determine conversion to AD from MCI over a 12-month period and found that a functional assessment questionnaire and Trails B accounted for 50% of the variance.

MCI is a clinical syndrome that is diagnosed when the underlying etiology and specific pathological processes that are contributing to the condition are unknown. However, research has attempted to identify conditions and mechanisms that result in the symptoms observed in MCI (e.g., memory, executive problems). In addition to AD (e.g., Sperling et al., 2011), others have also posited metabolic (e.g., diabetes; Craft, 2009), vascular (Villeneuve, Belleville, Massoud, Boti, & Gauthier, 2009; Zhang, Wei, Li, & Wang, 2011), and psychiatric etiologies (Qureshi et al., 2010; Zihl, Reppermund, Thum, & Unger, 2010), among others.

The National Institute on Aging and the Alzheimer’s Association working group recently developed specific criteria for the diagnosis of “MCI due to Alzheimer’s disease” (Albert et al., 2011). The au-
thers included clinical, cognitive, and functional criteria, as well as research criteria using biomarkers in an attempt to identify individuals in the pre-dementia stages of Alzheimer's disease, although again there is no known diagnostic test for Alzheimer's disease.

Despite the increase in research on MCI over the past decade, MCI as a useful and valid clinical entity continues to be a hotly debated topic (e.g., Allegri et al., 2008; Rockwood, Chertkow, & Feldman, 2007). Unfortunately, the neuropsychological diagnostic procedures are not consistent from study to study, or clinic to clinic, which has affected the reported prevalence and incident rates, rates of progression, and characterization of the syndrome (Dubois & Albert, 2004). This inconsistency has lead to great variability in results and makes accurate prognosis and study comparisons difficult. Furthermore, given the heterogeneous underpinnings, various subtypes, and variable prognoses and patterns of progression (or stability or recovery), some claim it is contentious to claim that MCI is a separate entity that deserves its own diagnosis. Some have even argued that the diagnosis of MCI, even when divided into subtypes, “lacks utility, and that MCI is not a diagnostic or disease entity but instead a descriptor of the severity of cognitive impairment” (Tröster & Fields, 2008, p. 540).

Others have argued that MCI is simply preclinical AD, and therefore, should not be differentiat-ed. Even the concept of a transitional period prior to onset of dementia has been called into question. For example, Goldman, Price, Storandt, Grant, Rubin, and Morris (2001) demonstrated that AD neuropathology can be observed in individuals who do not exhibit cognitive or functional decline. Similarly, Markesbery, Schmitt, Kryscio, Davis, Smith, and Wekstein (2006) demonstrated that, upon autopsy, neuropathologic findings of those diagnosed with aMCI were more similar to early stage AD than controls, suggesting that aMCI is actually early AD. In support of such claims, Meyer, Huang, and Chowdhury (2007) also presented structural neuroimaging findings identifying MCI prodrome states for AD, VaD, and Parkinson-Lewy body dementias. Thus, whereas there is some evidence suggesting MCI as a transition state between normal aging and dementia, the fact that not all individuals with MCI progress (and
some recover), and the final outcome can be very diverse (i.e., frontotemporal dementia, VaD, AD, depression), one can again question the utility of MCI as a separate, diagnostic clinical entity (Gauthier & Touchon, 2005).

### 1.2 Screening for MCI

Given the heterogeneity in presentations and possible underlying etiologies, it is not surprising that it might be difficult to identify a thorough, yet brief, screener that accurately identifies MCI and has strong psychometric characteristics that can address the needs of clinicians who are under time and cost restrictions. Screeners constitute the first step to identification if there is a disease process occurring that would warrant further comprehensive evaluation. Lonie, Tierney, and Ebmeier (2009) reviewed studies evaluating 15 cognitive screening instruments in their ability to identify MCI. The administration times for each measure ranged from one to thirty minutes. Only four measures comprehensively evaluated most cognitive domains determined to be important to Lonie and colleagues, i.e., orientation, memory, language, attention, visuospatial/perceptual processing, and executive functioning. These included the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), the Addenbrooke’s Cognitive Examination—Revised (ACE-R; Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006), the Cambridge Cognitive Examination (CAMCOG; Roth et al., 1986; Roth, Huppert, Mountjoy, & Tym, 1998), and the Consortium to Establish a Registry for Alzheimer’s Disease Neuropsychological Battery (CERAD-NB; Morris, Mohs, Rogers, Fillenbaum, & Heyman, 1988). The four comprehensive screeners generally had good sensitivity and specificity. As the number of domains evaluated decreased, the sensitivity of other screeners was reduced, while the specificity generally remained adequate.

The ACE-R is a 100-point test in which the MMSE is embedded. It evaluates five domains: attention and orientation, memory, fluency, language, and visuospatial ability. It takes between 12 and 20 minutes to administer in a clinical setting (Mioshi et al., 2006). Mioshi and colleagues reported that the Cronbach’s alpha coefficient for the ACE-R was 0.80, which is considered good. Despite the fact that
Lonie et al. (2009) reported sensitivity and specificity values for identifying MCI, these values in the original article actually refer to the sensitivity and specificity values for identifying dementia, not MCI. Values for MCI were not provided in Mioshi’s study. Thus, sensitivity and specificity data are not available for identifying MCI in the test development sample. However, they did report significant differences among controls, MCI, and dementia patients, such that those with MCI performed worse than controls, but better than those diagnosed with dementia.

Larner (2007) demonstrated excellent sensitivity (>0.95) for identifying MCI in an unspecified clinical sample when using the recommended cutoff scores of 88 and 82; however, the specificity was less than optimal (<0.75). Using a modified cutoff score of 75 drastically improved the specificity (>0.9) in his sample without significantly reducing the sensitivity (>0.9). This suggests that a lower cutoff score may be preferred in an unspecified clinical sample, i.e., similar to Larner’s “Cognitive Function Clinic.”

In a comparison of the MoCA and the ACE-R (Ahmed, de Jager, & Wilcock, 2012), both measures exhibited stronger sensitivity (90%) and weaker specificity (67%) in their sample of community-dwelling older adults. Optimal cutoff scores were consistent with those specified by Mioshi and colleagues (2006). The area under the curve was 0.822 (p < .0001). In sum, most studies on the ACE-R report excellent sensitivity for MCI, but weaker specificity. It is possible that using a lower cutoff score may improve specificity in a general clinical setting.

The CAMCOG is a subset of the Cambridge Mental Disorders of the Elderly Examination (CAMDEX), which is a standardized interview to aid in the diagnosis of dementia. The CAMCOG takes approximately 20-40 minutes to administer (depending on the severity of deficits) and evaluates several cognitive domains, including orientation, attention, memory, calculation, language, abstraction, perception, and praxis. The total score is out of 105-107 points, depending on whether the original or revised version is used. Performance on this measure is negatively affected by older age and lower education,
thus the generalizability of the test and its initial cutoff scores is limited (Huppert, Brayne, Gill, Paykel, & Beardsall, 1995).

In a Brazilian sample of well-educated older adults (Nunes et al., 2008) the Brazilian version of the CAMCOG demonstrated adequate diagnostic accuracy (AUC = .83) for discriminating between healthy controls and those diagnosed with MCI. An optimal cutoff score of 96 yielded poor sensitivity (64%) and adequate specificity (88%). The same group conducted the analyses again in a larger sample (Diniz et al., 2008) and reported stronger sensitivities (78-85%) but weaker specificities (64-75%), depending on the subtype of MCI being compared to healthy controls. The AUCs varied from .74 to .84, with multi-domain MCI consistently having the strongest diagnostic values. With more participants, the same research group reported similar AUCs ranging from .71 to .84 when using the CAMCOG to discriminate between controls and MCI and MCI subtypes (Aprahamian, Diniz, Izbicki, Radanovic, Nunez, & Forlenza, 2011).

Marcos and colleagues (2006) utilized the CAMCOG to predict conversion from MCI to AD in a Spanish sample, the majority of which had less than 10 years of education, and demonstrated good diagnostic ability to predict conversion (AUC = .88) with good sensitivity (92%) but weak specificity (68%). Gallagher and colleagues (2010) utilized the CAMCOG to predict conversion from MCI to AD and demonstrated a fair diagnostic ability to predict conversion (AUC = .79) with adequate sensitivity (.80) but poor specificity (.68) in an Irish sample. Conde-Sala and colleagues (2012), on the other hand, found near-chance diagnostic accuracy for the CAMCOG total score in the prediction of conversion from MCI to AD over a 5-year period in a Spanish sample of mostly females (68.7%) with diverse educational experiences, ranging from illiterate/no formal education to more than eight years of education. The authors determined that education had a significant negative impact on all subscales, except for the Memory (learning) subscale. These findings suggest that the CAMCOG may not be appropriate as a screening
measure for diverse patients due to its inconsistent psychometric properties and negative influence of older age and lower/poor quality of education.

The CERAD-NB consists of seven commonly administered measures of memory, praxis, language, and general cognitive functioning (Animal Fluency, Modified Boston Naming Test, Mini-Mental Status Exam [MMSE], Word List Memory, Constructional Praxis, Word List Recall, and Word List Recognition). It takes approximately 20 to 25 minutes to administer. Morris and colleagues (1988) initially demonstrated the CERAD-NB’s reliability across time points, for controls and “cases”, and for each subtest, as well as the validity of the CERAD-NB for differentiating between healthy controls and those diagnosed with varying severity levels of AD. Morris and colleagues (1988) noted a clear limitation of their sample, in that it was almost entirely white, well-educated, and middle class. Unverzagt and colleagues (1996) demonstrated strong education effects on performance on the CERAD-NB in a sample of African American adults aged 65 years and older. Specifically, greater education was associated with better performance across measures. Education levels alone accounted for large levels of variance in performance on the MMSE (36%), the Modified Boston Naming Test (30%), and Animal Fluency (29%). Both education and age accounted for significant levels of variance for Delayed Recall (22% and 10% of variance, respectively) and Recognition (18% and 4% respectively). Higher education and younger age were associated with better performance. The only score significantly associated with age was “Savings,” which accounts for words remembered after a delay, while accounting for the number of words initially learned. Thus, the authors suggested stratifying the normative data by education for all scores, except for Savings. This study demonstrates that caution should be taken when using the CERAD-NP with individuals with low educational attainments and older age, particularly in African American samples. Lower education has a significant negative impact on performance across all subtests, except for the Savings score.
Chandler and colleagues (2005) devised a CERAD-NP total score, which excluded the MMSE subtest, and determined that it was better at differentiating normal controls (NC) from those with MCI than the MMSE, but the total score was not better at differentiating those with MCI from participants diagnosed with AD. The Word List Recall was just as effective as the CERAD total score at differentiating between NC and the MCI group, suggesting that it is not necessary to administer the lengthy multi-subtest battery, when the single subtest is just as effective at identifying MCI. Age, education, and gender were all significantly correlated with the total score, so a demographic correction regression formula was provided. Similar to the findings by Unverzagt et al. (1996), demographics were shown to impact performance on the CERAD and should be taken into account when using the CERAD in diverse populations.

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998) includes 11 subtests (four of which are delayed recall and recognition tasks) assessing several cognitive domains with five index scores (Immediate Memory, Visuospatial/Constructional, Language, Attention, and Delayed Memory) and a Total Scale score. It takes approximately 25 minutes to administer and requires use of only one stimulus book and the record form. There are four equivalent alternative forms, allowing for re-testing over time. The subject is not required to be literate to complete the test. A recent study by Karantzoulis, Novitski, Gold, and Randolph (2013) reported high diagnostic utility in discriminating individuals diagnosed with “MCI due to AD,” which means that they had an amnestic form of MCI, and healthy controls. Specifically, the areas under the curve for the Delayed Memory Index (DMI) and the Total Scale Index scores were 0.90 and 0.88, respectively. Sensitivity, specificity, positive predictive value and negative predictive value scores were provided for DMI and Total Scale Index cutoff scores of one standard deviation and 1.5 standard deviations. Sensitivity values ranged from 32.1–71.6%, with one standard deviation being a more sensitive cutoff score and the DMI score demonstrating better sensitivity than the Total Scale Index. Specificity values were comparable for the DMI and the Total Scale
Index, (ranging from 89.9-96.3%) with a 1.5 standard deviation cutoff score being more sensitive than the one standard deviation cutoff score. This study suggests that the RBANS may be a viable option for brief cognitive screening for MCI.

The most sensitive instrument for discriminating between MCI and NC was the MoCA. This measure can be administered in approximately 10-12 minutes and was developed specifically to serve as a cognitive screener for MCI. Item development was initially based on clinical intuition of the first author and underwent modifications over five years of clinical use on patients referred to a memory clinic and healthy controls. In their 2005 study, Nasreddine and colleagues compared the sensitivity and specificity of the MoCA to the Mini Mental Status Exam (MMSE; Folstein, Folstein, & McHugh, 1975) for detection of individuals diagnosed with AD, MCI, or no diagnosis, using a more comprehensive neuropsychological exam as the criterion standard for diagnosis. The investigators used MCI criteria that focused on memory loss, thus their study sample likely consisted primarily of individuals with single-domain amnestic MCI and multi-domain amnestic MCI, while non-amnestic MCI was not evaluated. An education effect was observed, so they corrected this by adding one point for participants with 12 years of education or less. The internal consistency (Cronbach’s alpha) for the standardized items was .83. The MoCA had excellent sensitivity for detecting MCI and AD using an optimal cutoff score of 26 (90% and 100%, respectively), whereas the MMSE had poor sensitivity (18% and 78%, respectively). Specificity was excellent for the MoCA, which accurately identified 100% of the healthy controls, while the MMSE also had very good specificity (87%). Positive and negative predictive abilities for the MoCA were also excellent for MCI (89% and 91%, respectively) and AD (89% and 100%, respectively).

Subsequent research has criticized the MoCA for poor specificity, however. This may vary by population and setting. For example, McLennan, Mathias, Brennan, and Stewart (2011) found that while sensitivity for amnestic and multi-domain MCI was excellent, specificity was inadequate (50% for a-MCI and 52% for multi-domain MCI) in a cardiovascular population, where the primary complaint is not
memory or cognitive deficits. Rossetti, Lacritz, Cullum, and Weiner (2011) conducted a population-based study with demographically diverse participants on performance on the MoCA and found that 66% of individuals performed below the recommended cutoff score of 26 points, suggesting that caution should be applied when implementing cutoff scores and demographic factors should be considered. Luis, Kee-gan, and Mullan (2009) reported that in a sample of community dwelling older adults in the Southeast the recommended cutoff score of 26 demonstrated excellent sensitivity for MCI (97%) but the specificity was inadequate (35%). A modified cutoff score of 23 on the MoCA yielded excellent sensitivity (96%) and specificity (95%).

More relevant to the current study, Whitney, Mossbarger, Herman, and Ibarra (2012) reported that the specificity among middle-aged veterans was poor (57%). Using a lower cutoff score of ≤23 yielded a great specificity (75%), but decreased sensitivity from 86% to 72%. Of note, the modified sensitivity level was still far greater than that of the MMSE in this sample. Furthermore, Waldron-Perrine and Axelrod (2012) found that the optimal cutoff score on the MoCA for their sample of diverse veterans in an urban area was even lower than previous recommended scores of <26 and <23. For their sample, optimal cutoff scores varied from ≤19 to 21 when predicting impairment on various neuropsychological tests, as defined by performance that is one standard deviation below the mean. Even in a sample of well-educated community-dwelling older adults in England (Ahmed, et al., 2012), the optimal cutoff score was 23.5—much lower than the suggested cutoff score of 26 by Nasreddine et al. (2005). Thus, while the MoCA has been shown to be more sensitive than other cognitive screening measures (particularly the MMSE), the suggested cutoff score of 26 may not be adequate for diverse samples and, particularly, for a veteran population.

The Mattis Dementia Rating Scale, Second Edition (DRS-2; Jurica, Leitten, & Mattis, 2001) is another cognitive screening test that evaluates five cognitive domains (attention, initiation/perseveration, construction, conceptualization, and memory) in a hierarchical fashion, such that successfully complet-
ing initial, more difficult items allows examiners to give examinees credit for subsequent items. As such, administration of the test takes approximately 10-15 minutes in healthy older adults and 30 to 45 minutes in severely impaired individuals. An alternate form was also developed (Schmidt, 2004) and shown to be reliable, demonstrate strong correlations with the original form, and scores did not significantly differ across versions of the test (Schmidt, Mattis, Adams, & Nestor, 2005). Matteau and colleagues have demonstrated that the DRS-2 can successfully identify a-MCI (Matteau, Simard, Jean, & Turgeon, 2008), but is not sensitive enough to discriminate between a-MCI and MCI in patients with Parkinson’s disease, who demonstrate more of a frontal subcortical pattern of performance, e.g., poor executive functioning, initiation, and perseveration (Matteau et al., 2011).

Whereas it may be important to identify sensitive screeners that differentiate among types of MCI to help determine etiology and course, as with many other studies, Matteau and colleagues’ studies focused on amnestic forms of MCI and MCI associated with a specific, known etiology (Parkinson’s disease). More research is needed to determine whether the DRS-2 is sensitive enough to discriminate MCI from other clinical diagnoses, such as depression. Other researchers have demonstrated the predictive ability of the DRS-2 and its relation to independent functioning. Greenaway, Duncan, Hanna, and Smith (2012) demonstrated that performance on the DRS-2 and specific subscales related to executive functions and memory could predict functioning on independent activities of daily living (IADLs). In sum, the DRS-2 may be a viable screening measure to identify MCI and predict future outcomes.

In summary, the available, more comprehensive cognitive screeners described above and those reviewed by Lonie et al. (2009) have been shown to have generally adequate sensitivity to discriminate between individuals with dementia and healthy controls, and between individuals with MCI and dementia, but there is still some difficulty when discriminating between MCI and healthy controls, as well as between MCI and other clinical disorders, such as depression. In fact, Lonie and colleagues clearly stated, “specificity values for MCI in relation to depression are universally absent” (p. 913).
Lonie and colleagues (2009) also reviewed non-comprehensive screeners, including the most frequently used measure in clinics, the Mini Mental Status Exam (MMSE). It takes an average of 10 minutes to administer and assesses orientation, memory, language, attention, and visuospatial/perceptual processing. The measure has been criticized for its emphasis on the assessment of orientation, and overreliance on verbal tests, which is a concern when using the test on individuals with poor language skills and/or low education (Nieuwenhuis-Mark, 2010). The literature is clear that the MMSE has poor sensitivity for identifying MCI and discriminating it from healthy controls and AD (Mitchell, 2009), and performance varies significantly as a function of age and education. Specifically, with increasing age, performance goes down and the scores are more variable. With regard to education, with increasing years of formal education, performance on the MMSE increases and the variability is reduced (Crum, Anthony, Bassett, & Folstein, 1993). Additional factors, such as socioeconomic status and/or culture, as determined by neighborhood (i.e., barrio, transitional, or suburban areas) were shown to significantly impact performance on the MMSE (Espino, Lichtenstein, Palmer, & Hazuda, 2001). Aside from its psychometric properties, the MMSE is no longer in the public domain and, thus, is a costly option.

1.3 Veterans: A Unique Population

Although the MoCA appears to be a strong candidate for screening MCI, the cutoff score may be too high for a veteran population (Whitney et al., 2012) and may suggest the need for alternative norms for use in this population. Morgan, Teal, Reddy, Ford, and Ashton (2005) eloquently explained the rationale for acknowledging veterans as a unique population—and more specifically, veterans that utilize the Veterans Affairs (VA) medical centers as their primary source of care. Briefly, veterans in general are older, more likely to be male, better educated, and earn higher salaries than the general public (Morgan et al., 2005; National Center for Veterans Analysis and Statistics, 2011a; National Center for Veterans Analysis and Statistics, 2011b); however, veterans who use the Veterans Health Administration (VHA) as their primary source of care constitute an even more specialized population. These individuals are more
likely to be older, poorer, unemployed or underemployed, African American, are more likely to report poorer physical and mental health, and more chronic conditions, and have a greater self-identity as a veteran than veterans who do not use the VA healthcare system or the general public (Agha et al., 2000; Harada et al., 2002; Kazis et al., 1998). Several of these characteristics, including age, poverty, and lower education, put veterans at increased risk for cognitive impairment (Pleis et al., 2010).

1.4 The Saint Louis University Mental Status Examination

Given the unique nature of such a population, it might be suggested that measurement needs in this group might be different than others, and tests and/or normative data designed and developed specifically for veterans are needed. The only cognitive screener to date that has been developed specifically for, and for, a veteran population is the Saint Louis University Mental Status Examination, or the SLUMS (Tariq, Tumosa, Chibnall, Perry, & Morley, 2006). The authors aimed to address limitations of the MMSE, (i.e., limits when assessing highly educated individuals, poor evaluation of executive functions, overreliance on orientation) and to evaluate efficiently for MCI. However, the authors diagnosed subjects with mild neurocognitive disorder (MNCD), which is a DSM-IV-TR diagnosis that utilizes research criteria, not more common criteria for the diagnosis of MCI (Petersen et al., 1999 or Winblad et al, 2004 criteria).

MNCD requires the presence of two or more cognitive impairments (i.e., memory, executive functioning, attention or processing speed, perceptual-motor abilities, and/or language) lasting at least two weeks. There must be evidence that the impairment is due to a neurological or general medical condition, and the cognitive decline must be measured objectively on neuropsychological or cognitive testing. The cognitive deficits must cause marked distress or impairment in social, occupational, or other important areas of functioning and represent a decline from a previous level of functioning. Finally, the cognitive disturbance cannot meet criteria for a delirium, a dementia, or an amnestic disorder, and cannot be better accounted for by another mental disorder. This differs significantly from the criteria for
MCI, which only require one cognitive deficit, do not allow for significant impairments in daily functioning, and make no statement about it being better accounted for by another diagnosis. In fact, the accepted criteria for MCI provide no etiological explanation. No study to date has validated the SLUMS in a veteran sample diagnosed with MCI.

The SLUMS is an 11-item test that yields a score out of 30 possible points. It takes approximately seven minutes to administer and assesses orientation, attention (digit span), numeric calculation, immediate and delayed verbal recall, verbal fluency (animal naming), executive functions (clock drawing), figure recognition/size differentiation, and immediate recall of contextual verbal information (story). Tariq et al. (2006) did not describe in depth their process for inclusion or exclusion of items or their theoretical underpinnings. Certain items, such as the clock drawing, were included because they are observed to decline early in MNCD (Royall, Cordes, & Polk, 1998). It is possible that other items were included based on their ability to identify impairments in the domains delineated by the DSM-IV criteria for MNCD. MNCD requires impairment in two or more cognitive domains, thus essentially excluding single-domain subtypes of MCI. Furthermore, Tariq and colleagues did not report utilizing a comprehensive neuropsychological examination to diagnose MNCD, but instead diagnosed MNCD and dementia based on history, a physical and mental status examination, and laboratory findings. However, “evidence from neuropsychological testing or quantified cognitive assessment of an abnormality of decline in performance” (American Psychiatric Association, 2000) is required in the DSM-IV criteria for MNCD, so if a formal measure was used, it was not reported.

The final sample comprised 702 participants recruited prospectively from the Geriatric Research Education and Clinical Center (GRECC) at the Veterans’ Affairs Medical Center in St. Louis, MO. The mean age was 75.3 years old. The study sample was 62% cognitively “normal” (defined as not meeting DSM-IV criteria for MNCD or dementia), 26% were diagnosed with MNCD, and 12% were diagnosed with dementia. They were dichotomized into a group with less than a high school education (31%) and those
with a high school education or higher. It is unclear whether this dichotomization was due to performance being correlated with education, or preemptively to avoid potential education effects, as observed in the MMSE and the MoCA (Crum et al., 1993; Nasreddine et al., 2005). Based on Tariq and colleagues’ (2006) analyses, they delineated different cutoff score ranges for those with a high school education or more and those with less than a high school education.

Tariq and colleagues (2006) presented initial findings on the SLUMS and compared sensitivity and specificity measurements with the MMSE using receiver operating characteristic (ROC) curve analyses. Optimal cutoff scores for MNCD were identified for the SLUMS at 23.5 for individuals with less than a high school education and 25.5 for those with a high school education or higher. For dementia, cutoffs were identified at 19.5 and 21.5 for less than high school and high school or greater, respectively. Comparatively, cutoff scores were higher for the MMSE. For MNCD, cutoff scores were 28.5 and 29.5, and for dementia cutoff scores were 26.5 and 27.5 for individuals with less than a high school education and high school or higher, respectively. The area under the curve (AUC), a measure of overall diagnostic accuracy, was also estimated for the SLUMS and MMSE for MNCD subjects with less than a high school education. The SLUMS had a larger AUC compared to the MMSE (0.927 vs. 0.671) indicating greater diagnostic accuracy.\(^1\) Similarly, diagnostic accuracy was higher for the SLUMS than the MMSE in MNCD subjects with a high school education or above (0.941 vs. 0.643). Across education levels, animal naming, delayed recall, digit span, immediate and delayed recall, and clock hand placement significantly discriminated between controls and participants diagnosed with MNCD. However, it should be noted that the majority of items had AUC’s ranging from 0.50 to 0.59, with a few items with AUC scores in the 0.60 and 0.70 ranges. Although the analyses were statistically significant, likely due to the large sample size, the AUC values were in the “fail,” “poor,” and “fair” ranges.

\(^1\) For reference, an AUC value of one indicates perfect accuracy, whereas a value of 0.5 indicates random chance.
The SLUMS has a number of strengths and weaknesses. It is the only cognitive screener developed on and for a veteran population. It is also the only cognitive screener known that provides suggested cutoff scores for MNCD and dementia. It also takes into account potential education effects by providing different cutoff scores for those with less than a high school education and those with high school or more education. The instructions are standardized, allowing for ease of use by clinic staff. The SLUMS assesses rote memory, as well as contextual memory (paragraph) and addresses a number of domains or tasks that the MMSE does not, such as verbal fluency. Furthermore, the SLUMS is in the public domain, whereas the MMSE must be purchased.

The most important weakness of the SLUMS is the paucity of psychometric data regarding its validity and reliability. Tariq and colleagues (2006) did not provide a measure of reliability (i.e., Cronbach’s alpha, test-retest) for the SLUMS. It is crucial to establish that the exam is demonstrated to be a reliable measure of overall cognitive functioning. There are no studies within veteran samples comparing performance on the SLUMS to performance on well-validated neuropsychological measures that are known to be impacted in MCI, such as learning and memory or executive functioning; however, a recent study by Feliciano, Horning, Klebe, Anderson, Cornwell, and Davis (2013) evaluated the SLUMS in a community-based sample of adults aged 60 years and older. The authors noted that the distributions for both the SLUMS and the MMSE were negatively skewed, indicating that most participants performed well on the tests. The average SLUMS score was significantly lower than the MMSE, though (M = 23.85, SD = 4.42 and M = 28.50, SD = 2.64, respectively). They reported a significant correlation between the SLUMS and the MMSE (.75), and significant correlations between the SLUMS and each measure in a neuropsychological battery, which included the Trail Making Test (parts A and B), a list learning and recall test (Rey Auditory Verbal Learning Test), and a measure of executive functioning (Wisconsin Card Sort Test; WCST). Furthermore, the SLUMS significantly predicted each neuropsychological measure and explained a significant portion of the variance above and beyond demographics (age, sex, and education).
and the MMSE, with the exception of the number of errors on the WCST. This study is the first to demonstrate concurrent validity for the SLUMS with a neuropsychological battery; however, it does not examine the validity of the SLUMS in a veteran population or in a specialty (MCI) clinic sample, nor does it provide additional psychometric information, such as sensitivity, specificity, positive predictive values, or negative predictive values.

There are also no studies evaluating the ability of the SLUMS to discriminate between MCI and other clinical diagnoses, such as depression. The ability to provide differential diagnosis between a possible neurodegenerative cognitive impairment and that of depression has long been a goal in clinical services. In the elderly, in particular, cognitive dysfunction is often a feature of depression. Furthermore, cognitive changes may improve following treatment of depressive symptoms, but may not normalize (Steffens, 2008). Differential diagnosis is particularly difficult, given that the pattern of cognitive impairments in MCI and depression are very similar, consisting of deficits in attention, memory, and executive functions (Zihl et al., 2010). Dierckx, Engelborghs, De Raedt, De Deyn, and Ponjaert-Kristofferson (2007) reviewed the relationship between MCI and depression, noting that there are several hypotheses for the comorbid relation between depression and dementia, including (1) depression is an early stage of dementia, (2) depression is a risk factor for dementia, (3) depression is a reaction to the loss of cognitive and functional abilities associated with dementia, and (4) depression and dementia share common risk factors, such as cerebrovascular disease. There is no consensus on the nature of the relation between depression and dementia, but regardless, comorbid MCI and depression have been shown to increase the risk of converting to dementia (Penna, 2013).

Several studies have used the SLUMS, and it is increasingly cited in medical literature as an option for cognitive screening (e.g., Cordell et al., 2013; Galluzzi, Appelt, & Balin, 2010; Horvath, et al., 2011; Jak, Filoteo, Bondi, & Delis, 2012; Mansbach, MacDougall, & Rosenzweig, 2012; McPherson & Schoephoerster, 2012; Tamura & Yaffe, 2011). However, most studies that include the SLUMS either
used it as a predictive or outcome measure, therefore not examining its psychometric properties, or included such a small, heterogeneous sample that only descriptive statistics were provided when comparing the SLUMS to other cognitive screeners. This suggests the strong need for further evaluation of the SLUMS to determine its validity and reliability in other veteran populations. Brief summaries of studies that have evaluated the SLUMS or included it as a measure in their study are provided.

Brown, Lawson, McDaniel, and Wildman (2012) correlated performance on the SLUMS in a sample of 49 disability applicants (ages ranging from 20 to 64 with a mean of 44, education ranging from 5 to 18 years with a mean of 11) with performance on the Nevada Brief Cognitive Assessment Instrument (NBCAI). The NBCAI is a brief measure of estimated intelligence that correlates with the WAIS-III Verbal IQ. The authors found that it significantly correlated with the SLUMS in the disability sample, but did not significantly correlate in a sample of pre-surgical patients who had a higher average education level. It was argued that the reason for the lack of correlation in the latter comparison group was due to the minimal variability in performance on the measures.

A recent study by Buckingham et al. (2013) that was published in a non-peer reviewed university-based journal, compared older adults’ performance on the MMSE and the SLUMS. The participants were living independently, in assisted living, or in skilled nursing facilities. They completed both measures, and, consistent with the authors’ predictions, the participants performed more poorly on the SLUMS compared to the MMSE. This suggests that the SLUMS is a more challenging measure that may be more suitable to detect mild cognitive changes compared to the MMSE.

One study presented as a poster at the 2012 Alzheimer’s Association International Conference by the group that developed the SLUMS (Cummings-Vaughn, Cruz-Oliver, Malmstrom, Tumosa, & Morley, 2012) compared performance on the SLUMS to two other brief cognitive screeners, the MoCA and the Short Test of Mental Status (STMS) in 65 male veterans aged 60 years or older, with a median age of 78.5 years. Each measure was referenced against the Clinical Dementia Rating Scale (CDR), for which a
score of 0.5 suggests cognitive impairment. The authors determined that the SLUMS had comparable areas under the curve (ranging from 0.72 to 0.74), but the SLUMS demonstrated superior sensitivity (74.2%) compared to the MoCA (67.7%) and the STMS (61.3%). Specificity rates were not provided in the abstract. Thus, Cummings-Vaughn and colleagues suggested better sensitivity of the SLUMS for detecting cognitive impairment compared to the MoCA and the STMS.

Cruz-Oliver, Malmstrom, Allen, Tumosa, and Morley (2012) evaluated the SLUMS as a predictor of mortality and institutionalization over a period of seven to eight years in a veteran sample. The SLUMS was able to positively predict institutionalization and mortality in individuals identified as demented at baseline. The SLUMS also significantly predicted mortality rates, but not institutionalization, for participants identified as having MCI at baseline; however, the latter significant finding was found to be non-significant when adjusted for covariates. Thus, the SLUMS may serve as an alternative screening measure to the MMSE to predict mortality and institutionalization rates in individuals with dementia, but may not be helpful for prediction of outcome for individuals with MCI within 7.5 years. Members of the same group presented similar data at the Alzheimer’s Association International Conference in 2012 that included information regarding the rate of conversion from demented to MCI (28%) and from MCI back to normal cognition (21%; Cruz-Oliver, Malmstrom, Tumosa, & Morley, 2012). They noted that the most common explanations for these back-conversions were dementia therapy, hearing or vision correction, and discontinuation of anti-cholinergic medications.

The SLUMS was also used as a measure of cognitive impairment in a diverse sample of veterans with heart failure. Hawkins, Kilian, Firek, Kashner, Firek, and Silvet (2012) determined that 58% of their sample (n = 251) scored in the MNCD or dementia range on the SLUMS. Cognitive impairment was significantly associated with poorer medication adherence (70% for those scoring in the MNCD range and 73% for those scoring in the dementia range on the SLUMS compared to 78% medication adherence in non-impaired participants) and a number of demographic characteristics, including a positive association
with age, African American race, depression, and use of alcohol. Furthermore, participants who performed in the MNCD range or below on the SLUMS completed a neuropsychological evaluation. Performance on measures of verbal learning, immediate memory, and delayed verbal memory were the most impaired, as measured by scores 1.5 standard deviations below the normative mean. This study is important because it includes a veteran sample and validates the SLUMS for use in estimating medication adherence in a particular medical group (i.e., patients with heart failure); however, it did not correlate the SLUMS with the neuropsychological measures, provide psychometric properties of the SLUMS, or evaluate its ability to discriminate between various diagnostic groups.

Raji et al. (2005) evaluated a diverse sample of older adults for cognitive impairment who attended an eye clinic. It is important to note that this study used the SLUMS as an outcome measure. Additionally, they used very different cutoff scores than Tariq and colleagues suggested, for which the rationale and method are not explained. Specifically, Raji et al. considered a score of 20-27 for subjects with a high school education or above and 15-19 for less than high school education as meeting criteria for MCI, whereas Tariq et al. stated that scores of 21-26 and 20-24 were considered mildly impaired for persons with high school education or above and less than high school education, respectively. According to their cutoff scores, members of an ethnic minority (African American and Hispanic) and those with diabetes were more likely to be cognitively impaired. However, after controlling for diabetes, ethnicity was no longer a significant factor, suggesting that diabetes may impact performance on the SLUMS. Colberg, Somma, and Sechrist (2008) used the same modified SLUMS cutoff scores in their study evaluating cognitive functioning in individuals with type 2 diabetes. The authors determined that diabetes is not related to poor performance on the SLUMS, but performance was positively correlated with light to moderate exercise.

Stewart, O’Riley, Edelstein, and Gould (2012) conducted a pilot study comparing performance on the MMSE, SLUMS, and MoCA in a small sample of individuals in a rural long-term care facility. Partic-
Participants’ ages ranged from 48 to 89 years ($M=65.08$, $SD=8.83$, $n=40$) and they were diagnosed with either dementia or a psychiatric disorder (i.e., anxiety, depression, schizophrenia, schizoaffective disorder, bipolar disorder, deferred). No participant scored 30/30 on any cognitive screener. Those who met the cutoff score indicating cognitive impairment on the MMSE (24/30) also met the cutoff score on the MoCA (26/30) and performed in the “dementia” range on the SLUMS. Interestingly, six participants scored above the cutoff score for the MMSE, indicating non-impaired overall cognition, but scored in the “mild neurocognitive disorder” range on the SLUMS, and six other participants who were above the MMSE cutoff scored in the “dementia” range, suggesting that the MMSE may not be as sensitive to cognitive dysfunction and mild impairments as the SLUMS. This may be due to types of domains assessed (i.e., the SLUMS addresses logical memory and size differentiation, whereas the MMSE does not) or the sensitivity of the individual items. Nevertheless, the SLUMS correlated well with the MMSE ($r = 0.83$) and the MoCA ($r = 0.91$) in this sample.

Mossbarger, Whitney, Herman, and Mariner (2012) included the SLUMS as a measure to assess the rate of “carry-over” effects from brief cognitive screeners to a lengthier screening instrument, the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). The MMSE and the MoCA were also included. The authors reported 26% carry-over, defined as recalling a word used on the SLUMS recall task (e.g., “apple”) on the RBANS recall or recognition tasks. Thus, there were seven incidences of carry-over effects out of 27 administrations of the SLUMS+RBANS.

Duncan, Malmstrom, Grossberg, and Morley (2012) presented a poster in which they evaluated the Dr. Oz computer-based screening tool and used the SLUMS to classify their sample as having normal cognitive functioning, MCI, or dementia. Similarly, Aman and Thomas (2009) used performance on the SLUMS to confirm cognitive impairment for inclusion in a study on the effects of exercise in a sample of older adults with severe dementia. Guerreiro, Vaskov, Crews, Singleton, and Hardy (2009) used the
SLUMS to demonstrate cognitive decline over a period of four months in a case example of a woman with a gene mutation associated with Familial Fatal Insomnia.

The majority of studies cited above that use the SLUMS provided very limited psychometric information. They did provide some information regarding its validity or ability to discriminate between disorders, and how its results compared to other screeners. Furthermore, the only researchers that attempted to further evaluate the psychometric properties of the SLUMS were from the same group that developed the measure. There have been no independent evaluations. Additional psychometric evaluation and validation is needed. The current study proposes to evaluate the reliability and criterion, construct, and discriminant validity of the SLUMS in a veteran population and to compare its efficacy to that of a comprehensive neuropsychological battery, as well as to examine its ability to aid in differential diagnosis with depression. Additionally, the current study investigates the validity of the SLUMS in a veteran population in a different region of the United States, thus providing additional evaluation of the utility of this new cognitive screener that purports to identify MCI in veterans.

1.5 Proposed Study

The purpose of this study was to evaluate the SLUMS in a veteran population at the Atlanta Veterans Affairs Medical Center (VAMC). It improves upon previous research in a number of ways. First, the authors of the SLUMS argued that it is a good measure to identify MCI; however, the authors used MNCD as their diagnostic outcome. This study, on the other hand, utilized well-accepted criteria for MCI as an outcome measure. Additionally, a comprehensive neuropsychological battery was administered to support the MCI diagnosis, which is considered the gold standard for diagnosis (Cruz-Oliver, Malmstrom, Allen, et al., 2012; Cullen, O’Neill, Evans, Coen, & Lawlor, 2007). Furthermore, this study explored the utility of the SLUMS in a veteran population that is more racially diverse and located in a different region of the United States.
Several approaches to examine the validity of the SLUMS were implemented in the proposed study. First, construct validity was assessed by determining whether performance on the SLUMS is similar to performance on several other standardized neuropsychological measures that assess cognitive abilities known to deteriorate in MCI and dementia (convergent validity) and not similar to measures known to remain stable in MCI and early dementia (divergent validity). Specifically, convergent validity was assessed by comparing performance on the SLUMS with performance on the California Verbal Learning Test, Second Edition (CVLT-II), the Brief Visual Memory Test-Revised (BVMT-R), Digit Span, phonemic fluency (F, A, S), semantic fluency (animals), Similarities, and Trails B. These measures provide estimates of verbal and visual memory, auditory attention, verbal fluency, abstract reasoning, and executive functioning, which have all been demonstrated to deteriorate early in disease processes (Smith & Bondi, 2008). In addition, performance on the SLUMS was compared to performance on another cognitive screener, the DRS-2, for a subset of the sample (n = 51) that completed the measure. The DRS-2 is commonly used as an effective cognitive screener for dementia, and has demonstrated excellent discrimination of MCI in two studies (Matteau et al., 2008; Matteau et al., 2011), but takes longer to administer, requires additional materials, and is costly; thus, comparing the diagnostic ability of the SLUMS to the DRS-2 would be advantageous.

To assess divergent validity, performance on the SLUMS was compared to performance on the Boston Naming Test (BNT), Wechsler Test of Adult Reading (WTAR), Trails A, and Digit Symbol-Coding. These measures assess confrontation naming, word reading (an estimation of premorbid intelligence), and psychomotor processing speed—tasks that are not typically affected in MCI (e.g., Hannay, Howieson, Loring, Fischer, & Lezak, 2004, pg. 211).

Criterion validity was assessed by determining the diagnostic characteristics (e.g., sensitivity, specificity) of the SLUMS when differentiating between MCI and not MCI (i.e., anxiety, major depressive disorder, dementia, cognitive disorder, or no diagnosis); and between MCI and depression when com-
pared to 1) a comprehensive neuropsychological battery and 2) within a subset of patients with the DRS-2.

Differential diagnosis between MCI and dementia or depression has been one of the most difficult challenges for medical and mental health professionals working with older adults (Steffens, 2008).

The following hypotheses and study aims were proposed:

1. Overall performance on the SLUMS would be significantly correlated with performance on measures of memory (CVLT-II, BVMT-R), attention/working memory (Digit Span), verbal fluency (FAS, Animals), verbal abstract reasoning (Similarities), and executive functioning (Trails B), and, within a subset of the sample, with the DRS-2.

2. Performance on the SLUMS would not be significantly correlated with performance on confrontation naming (BNT), measures of estimated intelligence (WTAR), and processing speed (Trails A, Digit Symbol-Coding).

3. The criterion validity and diagnostic performance (i.e., AUC) of the SLUMS in participants with and without MCI (using both a “No Diagnosis” group and a “Non-MCI” clinical diagnosis group) will be evaluated. Additional diagnostic indicators (i.e., sensitivity, specificity, positive predictive values, negative predictive values) will also be evaluated. It is hypothesized that these values will be comparable to those from the comprehensive neuropsychological battery.

4. The SLUMS will show adequate discriminability between MCI and depression, and will show similar diagnostic ability when compared to those indicators from the comprehensive neuropsychological battery.

2 METHODS

2.1 Procedures

Archival data were used that were continuously collected from the Mild Cognitive Impairment (MCI) Clinic at the Atlanta VAMC. The Atlanta VAMC is located in a diverse, metropolitan area and serves
veterans from rural, suburban, and urban regions. Patients who were referred to the MCI Clinic reported subjective memory loss or cognitive dysfunction, or their healthcare provider or family member reported a decline in cognition. All potential patients completed a brief cognitive screener (i.e., MMSE, MoCA, or SLUMS). The sample consisted of participants seen between January 2010 and February 2013.

Results of the brief cognitive screener and all available medical histories and previous evaluations were reviewed for all patients prior to their scheduled visits. Each patient was interviewed and examined by a physician to determine risk factors and reversible sources of cognitive dysfunction, such as hypertension, diabetes, hypercholesterolemia, vitamin B deficiency, history of strokes or traumatic brain injury, sleep dysfunction, and history of substance abuse and mental health issues.

For the purpose of this study, the diagnoses of hypertension, diabetes, and hypercholesterolemia were based on expert panel guidelines (Chobanian et al., 2003; Expert Panel, 2001; The Expert Committee, 2003). Specifically, hypertension was defined as systolic BP >= 140 mm Hg, diastolic BP >= 90 mm Hg, or taking antihypertensive medication. Dyslipidemia was defined as a fasting total cholesterol level of 240 mg/dl or greater, or taking cholesterol lowering agents (e.g., statins). Diabetes was defined as a fasting plasma glucose level of 126 mg/dl or higher (7.0 mmol/L), taking insulin, or taking oral hypoglycemic agent. These diagnoses were considered important for consideration in this study due to their deleterious long-term impact on cognitive functioning (e.g., Arvanitakis, Wilson, Bienias, Evans, & Bennett, 2004; Goldstein, Ashley, Endeshaw, Hanfelt, Lah, & Levey, 2008; Li et al., 2011; Reitz, Tang, Manly, Mayeux, & Luchsinger, 2007). The majority of participants in the final sample (N = 148) were diagnosed with hypertension (n = 118, 78%) and/or dyslipidemia (n = 95, 62%), whereas about a third of the participants were diagnosed with diabetes (n = 52, 34%). The absence or presence of any of these conditions was not significantly related to any one diagnostic group (or “No Diagnosis” group) that was evaluated in the study.
Numerous medications are also known negatively to impact cognition in older adults. To document such medications and determine whether they significantly impacted performance in this sample, two independent lists were used to calculate a medication-related cognitive effect score. The first, the American Geriatrics Society 2012 Updated Beers Criteria for Potentially Inappropriate Medication (PIM) Use in Older Adults (AGS 2012 Beers Criteria), provides clinicians with medications deemed inappropriate for older adults, a rationale, recommendation (i.e., “avoid” or “avoid as a first-line antihypertensive”), the quality of the evidence, strength of the recommendation, and references to support their assertion. Any medications on this list noted potentially to cause cognitive changes in older adults were tallied for each participant to create a total score. A separate list, the Anticholinergic Cognitive Burden Scale (ACBL; available for download at www.indydiscovernetwork.org/anticholinergiccognitiveburdenscale.html; Boustani, Campbell, Munger, Maidment, & Fox, 2008; Campbell et al., 2009), was also used to create a total cognitive impact score. Neither cognitive impact score was significantly correlated with performance on the SLUMS (AGS Beers Criteria: \( r(148) = .11, p = .17 \); ACBL: \( r(148) = .05, p = .53 \)). It is possible that a significant relationship was not detected because of significant skewness in these index scores (and kurtosis for the ACBS score); however, transforming the variables did not improve the correlations. Furthermore, grouping the number of medications taken (i.e., 1 = no medications, 2 = one or two medications, 3 = three or more) and running Spearman rank order correlations also yielded non-significant results. It may be the case that medications that impact cognition do not have a significant effect until older ages, or that the SLUMS is not sensitive to medication effects.

In addition to a brief medical evaluation, a neuropsychological interview and assessment was completed. The assessment comprised a 30-minute semi-structured clinical interview with a licensed clinical neuropsychologist and a standard battery of neuropsychological assessments administered by a trained psychometrist or doctoral level trainee. The battery included an estimation of intellectual func-
tioning, individual and embedded measures of effort, and measures assessing language, processing speed, attention, verbal and visual memory, and executive functioning.

Effort testing was also included in the battery of tests. The Test of Memory Malingering (TOMM; Tombaugh, 1996) takes approximately 15 to 20 minutes to administer. The examinee was shown a series of 50 line drawings for three seconds at a time, then immediately shown 50 sets of two pictures and asked to specify which of the drawings they were just shown. There are two learning trials (if the examinee scores 45 or above, only one trial is required) and an optional retention trial following a delay. If a patient failed the effort evaluation, the decision to proceed with testing was at the discretion of the neuropsychologist. Only participants who earned a score of 45 or above in the final trial were included in the final analyses; thus, seven participants were excluded who exhibited suboptimal effort.

2.2 Participants

Specific inclusion and exclusion criteria are provided for clinicians who wish to refer their patients to the MCI Clinic; however, the clinic allows for a few exceptions (e.g., on rare occasions the patient has been prescribed memory enhancers, such as galantamine; some patients are older than 69 years old). The majority of patients referred met the following inclusion criteria: a) ages 50 – 69, b) subjective memory complaint from patient and/or memory complaints from family members, c) family member or caregiver able to attend appointment, and d) not previously diagnosed as having dementia. All of the following exclusion criteria also must have been met: they must not a) be on prescription memory enhancers at present, b) have severe hearing impairment, c) have severe visual impairment, d) be non-English speaking, e) suffer from active psychosis, f) engage in active substance abuse (within the past 2 months), and g) have no recent history of traumatic brain injury (within past 5 years). Two participants were included who carried a substance (alcohol) abuse diagnosis, but were not currently abusing (i.e., sober for the required minimum three months) and 25 participants had a history of significant past substance abuse noted in their medical charts. Thirty-five participants had a history of concussions,
and one participant suffered a moderate-severe head injury (all more than five years prior). Spearman rank coefficients demonstrated that history of past substance use and concussion history variables were not significantly correlated with performance on the SLUMS.

For the current study, all participants completed a SLUMS. All but eight participants had information regarding responses on individual items to be used in assessing internal reliability statistics (i.e., Cronbach’s alpha). In addition, with regard to neuropsychological testing, participants completed at minimum a) the DRS-2 or b) one measure from each domain (intelligence, attention, language, memory [visual and/or verbal], processing speed, and executive functioning). These measures were used by the clinical neuropsychologist to make a clinical diagnosis along with all other available patient clinical, laboratory and historical information. These neuropsychological measures were also used in this study to evaluate convergent and divergent validity of the SLUMS.

Participants who were given a clinical diagnosis other than MCI were also included to serve as the comparison group (i.e., “not-MCI”) to determine sensitivity and specificity of cognitive measures. Participants with depression as a primary diagnosis, who were not diagnosed with MCI, were also included to test the differential diagnostic utility of the SLUMS for MCI and depression, an important clinical decision.

Demographics of the final sample (N = 148) are summarized in Table 2. The final sample consisted of 148 participants with a mean age of 68.48 (range = 50-88, SD = 7.64). Females (n = 3) were excluded due to the extremely small number of subjects available; thus, the entire sample consisted of males. The level of education for this sample ranged from three years to 20 years (M = 13.24, SD = 2.82). Eighty two percent of the sample (n = 122) completed high school or higher levels of formal education. Those with a high school education and higher (M = 23.18, SD = 4.28) did not differ significantly in their performance on the SLUMS from those with less than a high school education (M = 22.04, SD = 4.68, t (146)= 1.31, p = .23) and education was not correlated with the SLUMS (r = .12, p = .19).
The sample was approximately two-thirds European American (n = 101, 68%) and a third African American (n = 47, 32%). Performance on the SLUMS differed significantly between European Americans ($M = 23.60, SD = 4.10$) and African Americans ($M = 21.64, SD = 4.63$, $t(146) = 2.60, p = 0.01$). This difference was still significant, even after controlling for education ($F(1, 144) = 4.01, p = 0.02$). African American participants, on average, performed more poorly on items 5 (Math) and 11 (Logical Memory) than European American participants, which accounted for the mean differences between the two groups.

A power analysis using G*Power (Erdfelder, Faul, & Buchner, 1996) to compute an estimated sample size adequate to detect a significant area under the curve (AUC; Wilcoxon-Mann-Whitney test) using a large effect size of $d = 2.2$ (as computed based on values from Tariq et al., 2006), $\alpha = .05$, and power = 0.95 suggested the need for a total sample size of 12 participants. Thus, this study has a large enough sample to conduct several ROC curve analyses. As the suggested sample size is very small, the calculated effect size of the current sample for the SLUMS (MCI vs. No Dx = 1.26 and MCI vs. MDD = 1.04) was used to validate the sample size used. Using G* Power, the suggested sample sizes were 32 and 44, respectively, suggesting that the current sample size is adequate for the current analyses.

2.3 Measures

The following measures were utilized by the VA clinicians to make a clinical diagnosis and are included in statistical analyses. Measures are organized according to common conceptual cognitive domains: cognitive screeners, estimated intelligence, attention, processing speed, language, memory, visuospatial skills, executive functioning. All scores included in the statistical analyses were converted to the normative sample scores to standardized values across measures (i.e., a T-score of 40 and a standard score of 85 are equal to a z-score of -1).


2.3.1 Cognitive screeners

The St. Louis University Mental Status Examination (SLUMS) was described above. Briefly, it is an 11-item cognitive screener purported to be sensitive to MNCD. It was developed using a veteran sample and is in the public domain, making it an attractive choice to medical and mental health professionals in the VA system. The SLUMS was not used to make a diagnosis so as to avoid circularity bias.

A subset of the participants (n = 49) were also administered the Mattis Dementia Rating Scale, Second Edition (DRS-2; Jurica et al., 2001). This measure was also described above. Briefly, it is a screener designed to identify dementia quickly and accurately in adults aged 56-105 years old. Depending on the level of impairment, it can take 10-45 minutes to administer. It assesses five domains (attention, initiation-perseveration, construction, conceptualization, and memory) and results in a total score. The DRS-2 has excellent validity properties (Matteau et al., 2011; Matteau et al., 2008). It uses the reliability from the original version, which has demonstrated excellent test-retest reliability (.97) and subscale correlation coefficients ranging from .61 to .94.

2.3.2 Estimated intelligence

The Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) was administered to estimate premorbid intellectual functioning. The WTAR is a word-reading test that includes 50 irregular words. It takes approximately 5-10 minutes to administer and includes normative data for adults aged 16-89 years old. It is standardized on a nationally representative sample. The WTAR was co-normed with the WAIS-III and provides estimates of WAIS-III IQ scores. The Estimated FSIQ score will be used in analyses to determine discriminant validity. Internal consistency across the age groups range from .90 to .97 for the U.S. sample.

2.3.3 Attention

The Digit Span subtest of the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV; Wechsler, 2008) was included as a measure of attention. The examinee is required to repeat strings of
numbers forward, backwards, and in sequence. The WAIS-IV is a well-validated test with nationally stratified normative data of individuals 16-90 years old. The Total Scaled Score will be used in analyses. The WAIS-IV manual reported adequate internal consistency for all subtests (> .70).

For some subjects, Trails A served as a measure of both attention and processing speed. This decision varied based on the neuropsychologist. Trails A (Reitan & Wolfson, 1985) is a visual-motor task that involves visual scanning, motor speed, and sequencing skills. Examinees are required to draw a line in numerical order from number 1 through number 25. The outcome for this task is based on the number of seconds needed to complete the task. Test-retest reliability has been shown to vary widely, ranging from $r = .36$ to $.79$ on Trails A (Bornstein, Baker, & Douglass, 1987; Dikmen, Heaton, Grant, & Temkin, 1999; Matarazzo, Wiens, Matarazzo, & Goldstein, 1974).

2.3.4 Processing speed

The Digit Symbol-Coding subtest of the WAIS-IV was included as a measure of processing speed. Examinees use a key at the top of the protocol with nine digit-symbol pairs to draw the corresponding symbol underneath each digit within a two-minute time limit. The score is based on the number of correct digit-symbol pairings completed in the time limit. Trails A also serves as a measure of processing speed for most participants.

2.3.5 Language

The Controlled Oral Word Association Test (COWAT; Reitan & Wolfson, 1985) is a test of verbal fluency in which examinees say as many words as they can that begin with a particular letter (i.e., F, A, and S). They have 60 seconds to respond for each letter. Animal Fluency is a similar measure that provides a semantic structure. Examinees are asked to name as many animals as they can in a 60-second time period. Performance on these measures is based on the total number of words generated, which
yields a final score that will be used in statistical analyses. Strong internal consistency has been demonstrated for the COWAT ($r = .83$; Ruff, Light, & Parker, 1996).

The Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 2001) is a confrontation naming task with 60 line-drawn objects that a person must name. If the examinee is unable to name the item or does not recognize the drawing, semantic and/or phonemic cues are provided. The total number correct is based on the number of spontaneously named items plus the number of items named with a semantic cue. The Boston Naming Test has been shown to be a reliable instrument (Spreen & Strauss, 1998).

### 2.3.6 Memory

Verbal learning and memory is assessed using the California Verbal Learning Test, Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000). A list of 16 words that can be organized into four different categories is recited over five trials. The examinee is to recall as many words as they can after each presentation. A distractor list is recited after the fifth trial, then the examinee is asked to recall as many words as they can from the initial list (short delay free recall). After the recall, they are cued (short delay cued recall). Following a 20-minute delay, examinees are asked to recall as many words from the initial list (long delay free recall), then they are cued again (long delay cued recall). A recognition task requires them to discriminate between words on the original list, words on the distractor list, semantically related words, and unrelated words. The Long Delay Free Recall score will be included in statistical analyses. Test-retest reliability was demonstrated to be strong (.80-.84; Woods, Delis, Scott, Kramer, & Holdnack, 2006).

The Brief Visual Memory Test, Revised (BVMT-R; Benedict, 1997) is used to assess visual learning and memory. Six designs are presented for 10 seconds over three trials. Following a 25-minute delay, the examinee is to recall as many figures as they can. They then complete a recognition task and an optional copy trial. The task is scored based on accuracy and correct placement of the drawings. The De-
layed Recall Score will be used in analyses. Alternate form reliability was demonstrated by Benedict, Schretlen, Groninger, Dobraski, & Shpritz (1996).

### 2.3.7 Executive functioning

The Similarities subtest of the WAIS-IV is a measure of abstract verbal reasoning. Participants are asked to describe how two things are alike. The items become more abstract as one proceeds through the task.

Trails B (Reitan & Wolfson, 1985) is a measure of cognitive set shifting. Examinees are asked to connect numbers to letters in numerical and alphabetical order, switching between numbers and letters. The outcome measure is the length of time it takes to complete. Test-retest reliability ranges widely from \( r = .44 \) to .89 (Bornstein, Baker, & Douglass, 1987; Dikmen, Heaton, Grant, & Temkin, 1999; Matarazzo, Wiens, Matarazzo, & Goldstein, 1974).

### 2.3.8 Emotional functioning

The Beck Depression Inventory, Second Edition (Beck, Steer, & Brown, 1996) is a 21-item self-administered questionnaire assessing depression symptoms. It takes approximately five minutes to complete and was developed for use in individuals aged 13 through 80 years. For each item, there are four statements of increasing severity of depression symptoms. The highest score is 63. The Cronbach’s alpha of the BDI-II is .93 (Beck et al., 1996).

### 2.4 Confirmatory Factor Analysis of the Neuropsychological Battery

A confirmatory factor analysis (CFA) was conducted to evaluate the conceptual measurement model that was thought to underlie the actual neuropsychological measures used by the VA. As with most neuropsychological measures, the tests included in the current battery do not necessarily measure only one isolated cognitive ability, but rather require several cognitive skills to complete the various tasks.
The initial CFA model specified five constructs or factors thought to underlie the available measures: attention, processing speed, language, memory, and executive functions. As described above, specific measures were hypothesized to represent these factors (see Figure 1; error variances, which are assumed to be uncorrelated between tests, are not shown for purposes of simplification). Model fit indices indicated a poor fit to this original model using the actual data collected in this sample of veterans. Analysis of the modification indices were investigated to guide re-specification of the model. A simplified three-factor model better explained the data collected and included the executive functions, language, and memory constructs (Figure 2). The semantic fluency measure (“Animals”) and Trails A measures were not included in this final model as they did not clearly load on any of the factors.

The final 3-factor model represents a good fit to the data. The chi-square test of model fit was \( \chi^2(17) = 18.78, p = .34 \), the RMSEA = .03, 90% CI [0.00, 0.08], \( p = .72 \), the CFI = .99, and the SRMR = .04. The factor loadings were also all significant at the \( p \leq 0.001 \) level, which also included significant correlations among the latent variables, as was expected. Individual measures accounted for a significant amount of variance in the various constructs, ranging from 31% (Similarities) to 57% (Digit Span).

2.5 Classification of Subjects

A diagnosis of MCI was developed in this study by the VA clinicians based on the criteria set forth by Winblad et al. (2004). Specifically, the criteria required that an individual (1) was not normal, and not demented (i.e., does not meet criteria for a dementing disorder according to DSM-IV or ICD 10), (2) preserved basic activities of daily living (ADLs; e.g., grooming oneself, dressing oneself)/but minimal impairment in complex instrumental functions (e.g., balancing a checkbook, cooking meals, driving, doing laundry, etc.), and (3) an impairment on objective cognitive tasks is present as reported by self and/or informant, and/or there is evidence of decline over time on objective cognitive tasks.

As this study used clinical and archival data that had already been collected, strict quantitative cutoff scores were not always utilized by the clinicians to determine cognitive deficits; however, they
reported that all impairments were at least one standard deviation below the normative mean. Functional abilities were determined by a clinical interview with the patient and an informant and/or reports by other healthcare professionals.

Dementia was diagnosed according to the definition in the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition, Text Revision (DSM-IV-TR; American Psychiatric Association, 2000) or the NINCDS-ADRDA criteria (McKhann, Drachman, Folstein, Katzman, Price, & Stadlan, 1984) (Table 1). Additional non-MCI diagnoses (i.e., anxiety disorders, Cognitive Disorder Not Otherwise Specified [NOS]) were made by the VA clinical neuropsychologist in accordance with the DSM-IV-TR.

Histories of mood and anxiety were available. Patients were categorized as having a current primary mood or anxiety disorder (i.e., major depressive disorder, generalized anxiety disorder, or post-traumatic stress disorder [PTSD]) for comparison purposes if 1) they were in a state of significant depression or anxiety at the time of the neuropsychological evaluation; and 2) met criteria for a major depressive episode or anxiety disorder according to the DSM-IV-TR as determined by the licensed VA clinical neuropsychologists. Self-report data using either the short form of the Geriatric Depression Scale (GDS; Yesavage et al., 1982; Sheikh & Yesavage, 1986) or the Beck Depression Inventory, Second Edition (BDI-II; Beck et al., 1996) were available as part of this diagnostic process. Classifications of anxiety or PTSD were made primarily by clinical interview of the participant and any available informants and/or diagnoses noted in medical charts.

Patients classified as “Depressed” had significant enough depressive symptoms that a diagnosis of MCI was not possible to make until the mood symptomology was better controlled, or it was clear on objective testing that there were no cognitive deficits and their subjective complaint of cognitive difficulties were secondary to a mood disorder. Therefore, it is possible that a “Depressed” patient may also have had an underlying MCI, but the depression was the primary diagnosis thought to be impacting their functioning at the time of the evaluation. Similarly, a person diagnosed with MCI may also carry a diag-
nosis of depression, but the depression is not present at the time of the evaluation, or is not believed to be the cause of the cognitive decline, given the pattern of deficits.

For the purpose of parsimony, participants were only assigned one (primary) diagnosis that most defined their clinical and cognitive presentation at the time of the evaluation. This procedure is more consistent with clinical research than that of randomized-controlled studies that typically require subjects in which only one disorder is present. The fact that the participants in the current study may have more than one diagnosis—even though only one is considered primary and contributory—is a limitation.

Finally, patients who were evaluated by the MCI Clinic due to memory concerns, but were found not to have a diagnosable condition, were classified as “No Diagnosis.” This means that they or their family members have concerns regarding changes in cognition or functioning, but their performance on neurocognitive measures and clinical and medical histories did not demonstrate sufficient evidence to warrant a DSM-IV-TR diagnosis.

Seventy-five participants were diagnosed with MCI (50.7%), 13 with dementia (8.8%), 8 with Cognitive Disorder Not Otherwise Specified (NOS) (5.4%), three with an anxiety disorder (2%), 23 with Major Depressive Disorder (MDD; 15.5%), and 26 were not given a diagnosis (16.9%). Due to small sample sizes for those diagnosed with dementia, Cognitive Disorder NOS, and anxiety, these individuals were analyzed in one very heterogeneous “Other” group. For select analyses, those diagnosed with MDD were included in the “Other” group to make sample sizes more comparable. Inclusion or exclusion of MDD in the “Other” group will always be specified.

The diagnostic groups differed significantly in age ($F(3, 143) = 4.23, p = .007$), such that those with MDD ($M = 64, SD = 5.07, 95\% CI [61.77, 66.15]$) were significantly younger than those not given a diagnosis ($M = 70.44, SD = 7.94, 95\% CI [67.16, 73.72], p = .018$) and those diagnosed with MCI ($M = 69.60, SD = 7.63, 95\% CI [67.85, 71.35], p = .011$). When considered together as one “Other” group, par-
Participants diagnosed with dementia, anxiety, or Cognitive Disorder NOS ($M = 67.21$, $SD = 8.06$, 95% CI [63.80, 70.61], $p > 0.05$) did not differ significantly in age from the other groups. One’s level of education did not differ based on clinical diagnosis ($F(3, 143)= 0.34$, $p = 0.80$).

### 2.6 Analyses

The statistical analyses were performed using the SPSS software version 19.0. Measurement of the internal consistency of the SLUMS was computed using Cronbach’s alpha and analysis of item characteristics. Associations between the SLUMS and the CVLT-II long delay free recall, BVMT-R delayed recall, Digit Span, phonemic and semantic fluency, Similarities, Trails B, WTAR, BNT, Trails A, and Digit Symbol-Coding were analyzed with Pearson correlations, and for a subset of participants, the SLUMS was also correlated with the DRS-2 total standard score. The following variables were not significantly correlated with the SLUMS scores: age; education; potential medication effects; histories of substance abuse, concussions, sleep disturbance, depression (not current primary MDD diagnosis), cardiac disease, and posttraumatic stress disorder. Family history of dementia and history of stroke negatively correlated with performance on the SLUMS (i.e., a significant history is associated with poorer performance), whereas a history of smoking and scores on the BDI-2 both positively correlated with the SLUMS.

The diagnostic utility of the SLUMS was evaluated using receiver operator characteristics (ROC) curve analyses. ROC curves were used to test the criterion validity and diagnostic performance of the SLUMS, the comprehensive neuropsychological battery, and the DRS-2. A curve was plotted based on the sensitivity and 1-specificity (false positives) for every value of the SLUMS. From these values, an optimal cutoff was determined for the sample that maximized sensitivity and specificity. The area under the curve (AUC) was calculated to estimate the overall diagnostic utility of the SLUMS. Positive predictive values (PPV; the proportion of positive test results that are true positives) and negative predictive values (NPV; the proportion of subjects with a negative test result who do not have the target disorder) were also calculated. ROC curves, AUC, PPV and NPP were also calculated for the comprehensive neuro-
psychological battery and the DRS-2 to determine how the SLUMS compared to another cognitive screening measure (DRS-2) and the gold standard, a comprehensive neuropsychological battery, when discriminating between persons with MCI and those without MCI.

The results of the comprehensive neuropsychological battery were represented two different ways in the ROC analyses: 1) as the number of tests impaired—defined as scores one standard deviation below the normative mean, and 2) as the number of tests impaired—defined as scores 1.5 standard deviations below the normative mean. This was to help clinicians determine in the future which cutoff level is most sensitive to MCI. Results demonstrated that using a 1.5 standard deviation did not improve the accuracy in discriminating between those with MCI and the other groups. Therefore, only the analyses using a one standard deviation cutoff are reported.

The diagnostic accuracy of the neuropsychological battery was also evaluated in terms of the underlying factors measured, as determined by the CFA results (i.e., three factors were identified: Executive Functions, Language, and Memory). First, the number of impaired measures that comprise an individual factor was used to discriminate between MCI and the other comparison groups. For example, the Memory factor is comprised of the CVLT-II Delayed Recall and the BVMT-R Delayed Recall scores. Thus, deficits in zero, one, or two measures from the Memory factor were used to discriminate between groups. In this way, it could be determined whether one factor was more sensitive than another, which helped elucidate the next step in the analyses.

Following the previous analyses, an average ‘factor’ score was computed for each observed measure for each factor (e.g., averaging the z-scores for the CVLT-II and BVMT-R Delayed Recall scores for the Memory factor). A score was considered a deficit if it was negative one z-score or below. An ROC curve analysis was conducted to determine the optimal number of deficits among the three factors for discriminating between MCI and the other groups.
3 RESULTS

3.1 Data Screening

Standardized test scores for each variable were converted to z-scores in order to report values in a unified metric. The data were examined for univariate and multivariate outliers, skewness, kurtosis, linearity, and multicollinearity and all were determined to be adequate. Extreme scores (i.e., more than three standard deviations above or below the mean) were recoded to equal a z-score of +/-3 (n = 9; Tabachnick & Fidell, 2007, pp. 46-47). The criterion for multivariate outliers was Mahalanobis distance at \( p < .001 \). There was no evidence of extreme multivariate outliers. Skewness and kurtosis were explored utilizing statistical and graphical (histograms) inspection techniques. No concerns were noted.

The average SLUMS score for the current sample was 23, and the scores ranged from 10 to 30 (SD = 4.36). A one-way analysis of variance (ANOVA) was conducted to determine whether SLUMS performance differed significantly amongst individuals with no diagnosis, MCI, MDD, or another clinical diagnosis. Two different measures of homogeneity of variance differed, so the results should be interpreted with caution. Specifically, the Levene statistic was significant (\( W = 4.29, p = .006 \)); however, the \( F_{\text{max}} \) test was well within normal limits. As might be expected due to the heterogeneity in etiologies, the “Other” group had the largest variance. Group sizes might also be impacting the results, as the largest group (MCI) is a little over three times the size of the smallest (MDD); however, this still meets the rule-of-thumb that sample sizes should not exceed a ratio of 4:1 (Tabachnick & Fidell, 2007, p. 88).

The ANOVA revealed significant differences in performance amongst the groups (\( F(3, 143) = 25.54, p < 0.001 \)). Post-hoc Bonferonni comparisons demonstrated that individuals diagnosed with MCI (\( M = 22.43, SD = 3.32 \)) performed significantly better than those with another clinical diagnosis (i.e., anxiety, Cognitive Disorder NOS, or dementia; \( M = 18.50, SD = 4.93; p < 0.001, 95\% \text{ CI} [1.69, 6.16], \) Cohen’s \( d = 0.94 \)), and significantly poorer than those diagnosed with MDD (\( p < 0.001, 95\% \text{ CI} [-5.80, -1.26], \) Cohen’s \( d = 1.05 \)) and those with no diagnosis (\( p < 0.001, 95\% \text{ CI} [-6.02, -1.61], \) Cohen’s \( d = 1.26 \)). Interes-
ingly, those with no diagnosis ($M = 26.24, SD = 2.70$) and those diagnosed with MDD ($M = 26, SD = 3.50$) did not perform significantly differently ($p = 1.00$, 95% CI [-2.47, 3.04], Cohen’s $d = 0.09$).

Pearson product-moment correlations were conducted for SLUMS scores, demographic variables, and neuropsychological test variables (Table 3). The SLUMS total score significantly correlated with the executive functions ($r(139) = 0.32, p < 0.01$), language ($r(143) = 0.44, p < 0.01$), and memory ($r(142) = 0.50, p < 0.01$) factor scores (i.e., the averaged $z$-score of each test that compose the factors). The SLUMS total score was also significantly correlated with every neuropsychological measure (including the DRS-2; $p < 0.01$), except for Trails B ($r(125) = 0.13, p = 0.15$). The SLUMS also significantly correlate with the BDI-2 ($r(53) = 0.32, p = 0.02$).

### 3.2 Reliability

Although screening measures are designed to cover many conceptual areas and not provide measurement of a single construct, one would expect that in the case of a broad and general cognitive decline, different items, measuring different cognitive abilities, would be somewhat related. If early cognitive decline only impacts specific areas of cognitive functioning, and not global systems, then the item relationships may be less related. When using all available data ($n = 140$), the SLUMS appeared to have poor internal consistency, with $\alpha = 0.57$. Item three, “What state are we in?” was not included, as everyone answered this question correctly. Inter-item correlations (Table 4) were low, with only four correlations reaching the 0.3 level. Three items demonstrated low correlations with the total score (Items 1, 2, and 10: Day, Year, and the Triangle item, respectively). The Cronbach’s alpha increased marginally if the individual item was deleted. Removing all three items increased alpha, but only to .59. Digit Span (repeating numbers backward) correlated the strongest with the overall SLUMS score, suggesting that it may be the best single indicator of performance on the SLUMS.
3.3 ROC Curve Analysis for the SLUMS

ROC curve analyses were conducted to determine the diagnostic accuracy of the SLUMS in the
detection of MCI compared to participants with a) no diagnosis, b) non-MCI diagnoses (includes MDD),
and c) a diagnosis of major depressive disorder (in the absence of a diagnosis of MCI). Sensitivity and
specificity values are provided for all cutoff scores in Table 7.

3.3.1 MCI vs. No Diagnosis

The area under the curve (AUC) discriminating between those diagnosed with MCI and no diag-
nosis was .82 with standard error (SE) = 0.05 and 95% CI [.72, .91] (p < .001). The optimal cutoff score for
differentiating MCI from patients with no diagnosis was slightly lower than the cutoff score reported by
Tariq et al. (2006). Specifically, a score of 25 or less was suggestive of a diagnosis of MCI (Sensitivity =
81%, Specificity = 68%). The positive predictive value (PPV) was 88% and the negative predictive value
(NPV) was 55%.

3.3.2 MCI vs. MDD

The AUC for discriminating between those diagnosed with MCI versus MDD was .79 (SE = 0.06, p
< .001, 95% CI [.68, .90]. The optimal cutoff score for identifying MCI was 24 (Sensitivity = 73%, Specifici-
ty = 70%, PPV = 89%, NPV = 44%).

3.3.3 MCI vs. Other Diagnosis

The AUC for discriminating between those diagnosed with MCI versus another diagnosis was .51
(SE = 0.06, p = .83, 95% CI [.40, .63]. The optimal cutoff score for identifying MCI was 24 (Sensitivity =
73%, Specificity = 40%, PPV = 73%, NPV = 49%).

Overall, the SLUMS is best at discriminating between those with MCI and those with either no
diagnosis or MDD. Its low rates of negative predictive value compared to the positive predictive value
suggest that a score above 24 or 25 does not necessarily indicate that an individual does not meet criteria for MCI.

### 3.4 ROC Curve Analysis for the Neuropsychological Battery

ROC curve analyses were conducted to determine the diagnostic accuracy of the broader neuropsychological battery (i.e., the number of impaired measures) in the detection of MCI compared to participants with a) no diagnosis, b) a diagnosis of major depressive disorder (in the absence of a diagnosis of MCI), and c) non-MCI diagnoses. This would conceptually set the upper limit of the range of possible success for the SLUMS, as the neuropsychological battery is considered to be more definitive and comprehensive for making such a diagnosis. Sensitivity and specificity values for the cutoff scores are provided in Tables 8.

#### 3.4.1 MCI vs. No Diagnosis

The AUC for discriminating between those diagnosed with MCI and no diagnosis was .89 with $SE = 0.04$ and 95% CI [.81, .96] ($p < .001$). Three deficits resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, three deficits resulted in a sensitivity of 79%, specificity of 88%, PPV of 95% and NPV of 59%.

#### 3.4.2 MCI vs. MDD

The AUC for discriminating between those diagnosed with MCI versus MDD was .79 ($SE = .06$, $p < .001$, 95% CI [.68, .91]. Three deficits resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, three deficits resulted in a sensitivity of 79%, specificity of 71%, PPV of 90%, and NPV of 48%.

#### 3.4.3 MCI vs. Other Diagnosis

The AUC for discriminating between those diagnosed with MCI versus another diagnosis was .59 ($SE = 0.06$, $p = .10$, 95% CI [.48, .70]. Four deficits resulted in the most optimal diagnostic indicators for
identifying MCI. Specifically, four deficits resulted in a sensitivity of 56%, specificity of 56%, PPV of 67%, and a NPV of 44%.

As observed when using the SLUMS to discriminate participants with MCI from other clinical and non-clinical groups, the neuropsychological battery is more efficient at discriminating individuals with MCI from those without a diagnosis or those with MDD than from participants with another clinical diagnosis. The one standard deviation deficit criterion of three or more deficits is more effective at correctly identifying MCI (PPV ranging from 90% to 95%) than correctly identifying no diagnosis (NPV = 59%) or a diagnosis of MDD (NPV = 48%).

3.5 ROC Curve Analysis for the Factor Scores

ROC curves were also conducted with the factor scores as described previously. Please refer to Table 9 for the results and to Appendix B for a more detailed description of these analyses.

In brief, essentially no measure that was evaluated in this study would be considered adequate, on its own, to discriminate between MCI and other clinical disorders in a screening situation with veterans. There was, however, some evidence in support of using the SLUMS and neuropsychological assessments to discriminate between MCI and no diagnosis or MDD. Effect sizes (Table 10) across measures (SLUMS, neuropsychological testing, and factor scores) were generally large when discriminating between individuals diagnosed with MCI and those with no diagnosis or MDD (ranged from 1.07 to 1.65), with the exception of the Executive Functions factor (effect size for MCI vs. MDD was 0.67). The effect sizes for discriminating between MCI and other clinical diagnoses were smaller, ranging from 0.06 for the SLUMS to a medium effect size (0.40) for the Memory factor score. Interestingly, when those with MDD were removed from the Other Diagnosis group, the effect size increased to a large 0.93. Effect sizes for the relationship between MCI and No Diagnosis were consistently large and stronger than the other relationships, although MCI versus MDD was also large across measures. The neuropsychological battery (measured as the number of impaired measures) had the largest effect sizes, with the excep-
tion of MCI versus Other Diagnosis (small effect size), which had a larger (medium) effect size for the Memory factor. It must be noted that the effect sizes, AUCs, and diagnostic indicators for the SLUMS were comparable to those of the comprehensive neuropsychological battery, with only the specificity values being markedly lower for the SLUMS when discriminating between MCI and no diagnosis.

4 DISCUSSION

The current study was designed to permit evaluation of the psychometric properties of the SLUMS in a broad sample of veterans referred to a MCI clinic for evaluation. As might be expected, the sample consisted primarily of individuals with MCI. It is important that the SLUMS validly be able to screen for individuals who may meet the criteria for MCI and those who do not meet the criteria. It is also beneficial if it can discriminate between MCI and another cognitive or affective disorder, in addition to non-cognitively impaired adults. Furthermore, given the cost and lengthy time requirements of comprehensive neuropsychological assessments, the diagnostic accuracy of this brief cognitive screener needed to be evaluated.

According to Shulman (2000), an effective screener should be quick to administer; well-tolerated and accepted by the patients; easy to score; relatively independent of culture, language and education; have good reliability; have high levels of sensitivity and specificity; have concurrent validity (a type of criterion validity); and have predictive validity. A screener is also not an appropriate tool with which to assign a diagnosis, but rather it should alert clinicians to the need for further and more comprehensive evaluation. Based on past report by Tariq et al. (2006), the SLUMS takes an average of seven minutes (± three minutes) to administer. It is simple and fast to score—simply add the item scores (10 scored items). The SLUMS may not be free from the effect of culture, language, or education, which is a weakness for use in diverse populations, but there is an adjustment for low education. One type of predictive validity was demonstrated by Cruz-Oliver, Malmstrom, Allen, and colleagues (2012), in that the
SLUMS predicts mortality and institutionalization in older adults; however, after controlling for covariates, predictive validity for those performing in the “MCI” (or “MNCD”) range was no longer statistically significant, whereas the results remained significant for those performing in the “dementia” range. Thus, more research is needed to fulfill the criterion of adequate predictive validity.

The results of the current study demonstrated poor internal consistency for the SLUMS. Nunnally (1978) discussed preferred reliability values in detail, suggesting that for applied purposes, such as a clinical test that may aid in diagnosis, a reliability of 0.90 is a minimum, and 0.95 is more preferred. The SLUMS does not meet these reliability criteria if one uses only the internal consistency measure evaluated in this study (α = 0.57). Whether test-retest reliability or other types of reliability would show different results is unknown at this time since this study provides the only currently available estimates of its reliability Cronbach’s alpha.

Although the SLUMS is purported to measure diverse cognitive domains, ranging from visuospatial skills to executive functions, performance on this measure is still considered reflective of a patient’s overall cognitive abilities. At the very least, one would assume that performance by individuals with no diagnosis might help improve the reliability score; however, the alpha value dropped from .57 to .43 when including only participants with no diagnosis in the analysis. However, the latter results of the Cronbach’s alpha should be interpreted with caution, as the subsample of individuals without a diagnosis with available item-level data from the SLUMS is considerably small (n = 21) and does not meet typical criteria to be considered a valid estimate of the population coefficient (Yurdugül, 2008). Interestingly, the Cronbach’s alpha within the MCI group only (n = 72) was .26, the dementia group (n = 13) was .43, and the MDD group (n = 22) was .26. While the latter analyses are qualitative, they serve as starting points to further research on reliability studies in larger, more clearly defined samples.

It is an unfortunate fact that the authors of many cognitive screening measures do not publish internal consistency values (Cullen et al., 2007). Some researchers will report the reliability statistics for
a particular sample of interest, which may vary drastically across settings and populations. For example, 
the Cronbach’s alpha for the MMSE has ranged from poor (.54) to excellent (.96) across studies (Tom-
baugh & McIntyre, 1992). Regardless, it is clear that even short screening measures of general cognitive 
functions, such as that demonstrated by the MMSE and the MoCA, are capable of exhibiting adequate 
internal reliability.

Nevertheless, without strong reliability, a measure cannot typically show strong validity in its 
measurement of a construct. In the case of the SLUMS, the construct it purports to measure—overall 
cognition—is, by nature, heterogeneous. It is possible, however, that if the SLUMS measures more than 
one construct, this may reduce the internal consistency of the measure. That being said, the existence of 
multidimensionality in such a short measure may be inconsequential for clinical interpretations. Specifi-
cally, a) it is very likely that any subscales/or individual items would be less reliable than the total score, 
b) the items within possible subscales may not be theoretically or clinically relevant, but statistically they 
correlate with one another, c) it is likely that the possible subscales would be significantly correlated 
with each other and with the total score, and, therefore, be redundant, and d) the construct that the 
subscale is measuring may be better measured by the more reliable total score (Reise, Bonifay, & 
Haviland, 2013).

It is also important to consider other forms of reliability that were not assessed here. Given that 
the SLUMS measures diverse cognitive abilities (that may together comprise a general factor), a split-
half reliability analysis may be less relevant. More informative reliability studies may include inter-rater 
reliability and test-retest reliability. It may be that due to the heterogeneous nature of a measure of 
“overall cognition,” Cronbach’s alpha values may differ, but performance on the test across time periods 
in cognitively intact (or non-neurodegenerative patients) should remain relatively the same. Further-
more, staff that administer the SLUMS should be doing so in such a way that they read instructions and 
score items in a consistent manner across examiners. In fact, this last point was not evaluated in the cur-
rent study or in the development of the MCI Clinic, which brings up issues of threats to reliability, including rater reliability (e.g., exam may not be administered in a standardize format, differences in rate of stimulus presentation, the effects of accents or rapport), situational reliability (e.g., noise, temperature, distractions), and subject reliability (e.g., fatigue, hunger, irritability). Thus, although the SLUMS appears to have poor internal consistency in the current sample, additional studies should evaluate inter-rater reliability, test-retest reliability, and conduct prospective research with trained staff administering the SLUMS to fully evaluate Shulman’s (2000) criterion of good reliability.

Within the context of poor internal consistency in the current sample, it does still appear as though the SLUMS is tapping into similar cognitive abilities as the comprehensive neuropsychological assessment, as the total score significantly correlates with all individual tests (with the exception of Trails B) and the executive functions, language, and memory factor scores, suggesting that the SLUMS may be a valid screening indicator of a patient’s performance on a more comprehensive neuropsychological battery. Interestingly, the SLUMS also correlated with measures of confrontation naming (BNT), estimated premorbid intellectual functioning (WTAR), and processing speed (Trails A, Digit Symbol-Coding), which did not support the second hypothesis of this study regarding its discriminant validity, but rather suggest that there may exist some shared attributes among these tasks.

Several ROC curve analyses were conducted to evaluate the diagnostic accuracy of the SLUMS, particularly compared to that of a comprehensive neuropsychological battery. Performance on the SLUMS and the neuropsychological examination was able to adequately discriminate between individuals diagnosed with MCI and those given no diagnosis or those diagnosed with MDD. The AUCs and 95% confidence intervals were comparable when using the SLUMS and the number of impaired neuropsychological measures. However, all measures examined consistently demonstrated inadequate diagnostic indicators when discriminating between those with MCI and those with non-MCI diagnoses (i.e., Cognitive Disorder NOS, anxiety, dementia, and depression). The most obvious explanation for this result is
that the non-MCI group reflected heterogeneous diagnoses, small sample sizes, and varying levels of severity. Due to small sample size, the non-MCI group was unable to be separated into more homogeneous groups, which would have allowed an expansion of these results. The non-MCI group included those with dementia, individuals suffering from multiple etiologies (e.g., sleep disturbances and/or sleep apnea, multiple medications that affect cognition, vascular risk factors, obesity), stroke survivors, major depressive disorder, and anxiety/posttraumatic stress disorder. Individuals with such diagnoses do not present with similar neuropsychological profiles, although their differentiation is an important problem for clinicians in such MCI clinic situations. Future studies should form more diagnostically homogeneous comparison groups to screen more effectively for MCI versus these other diagnoses.

Briefly, the current study also evaluated the underlying factor structure of the neuropsychological battery. Three factors were identified: Executive Functions, Language, and Memory. Individual ROC curve analyses were conducted to determine the accuracy with which each of these factors, independently, could differentiate between MCI and non-MCI (i.e., No Diagnosis, Other Diagnosis, or MDD). These analyses helped to begin to interpret the ROC curve analyses using a total factor score—one that counted the number of impaired individual factor scores (i.e., a score of one through three). By examining Table 10, one can see that an impairment in one factor resulted in the optimal indicators for identifying MCI in each analysis. Specifically, based on the classification statistics, it is likely that the Memory factor is implicated in best discriminating between MCI and no diagnosis or a non-MCI diagnosis, whereas the Language factor is implicated in best discriminating between MCI and MDD.

Thus far, the SLUMS meets a number of criteria set out by Shulman (2000) for a good screening measure. It is quick to administer; well-tolerated and accepted by the patients; easy to score; has good sensitivity and specificity; and has concurrent validity and predictive validity. More research is needed to examine aspects of reliability in addition to internal consistency.
As this is the first additional study known to examine the psychometric properties of the SLUMS in a veteran sample, it is important to discuss similarities and differences with the original publication of the SLUMS by Tariq and colleagues (2006). The authors of the SLUMS did not report a measure of reliability in their 2006 article, but did demonstrate excellent diagnostic abilities for identifying MCI and dementia compared to no diagnosis (AUC = .94, sensitivity = 95%, specificity = 76%). There may be several explanations for these study-specific differences, one of which is sample differences. Compared to the original SLUMS sample (Tariq et al., 2006), the current study sample is younger and a higher percentage completed high school (an average education level was not provided by Tariq et al., 2006). Tariq and colleagues did not describe the ethnic breakdown of their sample, but did report that it was primarily white, thus our sample is likely more diverse, despite being a majority white (67.1% European American and 32.2% African American). The finding that European Americans and African Americans differed significantly in their performance on the SLUMS highlights the importance of studying performance in ethnic minority groups and investigating possible sources of variability (e.g., socioeconomic status, quality of education, language or cultural differences). This difference persisted when education was held constant. Thus, in this case, race/ethnicity is not likely acting as a proxy for years of formal education, as is sometimes the case. However, controlling for the WTAR predicted FSIQ scores, a measure of word reading, accounted for the significant relation between ethnicity/race and SLUMS performance. Word reading skills may better reflect one’s quality of education than the self-reported number of formal years of education (Manly, Jacobs, Touradj, Small, & Stern, 2002).

Sample size may also affect the power of the statistical analyses. A power analysis was conducted prior to conducting this study that demonstrated adequate power for our sample size. Tariq et al. (2006) enrolled 705 participants, which is more than four times the sample size of the current study. Given the a priori power analysis, however, it seems unlikely that sample size would be the reason for such differences in diagnostic statistics (i.e., AUC, sensitivity, specificity) and poor internal consistency of
the measure. Furthermore, although the sample size of the current sample would not be large enough to detect a small effect size (a sample size of 290 would be required, at 80% power), the calculated effect sizes for all comparisons of MCI against those with no diagnosis or MDD were large. Thus, differences in sample size likely do not account for any discrepancies in study results.

In addition to demographic and sample size differences, the base rates of diagnoses in the study sample differed from that of the development sample (Tariq et al., 2006). In the original sample, more than half received no diagnosis (62.4%) compared to 17.1% of the current sample; a quarter of the original sample was diagnosed with MCI (25.5%) whereas 50% of the current sample was diagnosed with MCI; finally, 11.6% of the original SLUMS sample was diagnosed with dementia, whereas 8.6% of the current sample was diagnosed with dementia. Thus, it is most likely that differences in types of settings (i.e., a general clinic vs. a specialty MCI clinic) between the studies resulted in differing base rate of MCI, such that the current sample has a higher base rate. This would have a significant effect on and may explain the observed differences of the diagnostic indicators.

Interestingly, there is some evidence in the current sample that the SLUMS may be effective at discriminating adults with no cognitive impairment (no diagnosis) from more than simply MCI. The SLUMS was able to discriminate between those with no diagnosis and dementia and Cognitive Disorder, NOS quite well (AUC = .93, SE = 0.04, 95% CI [0.86, 1.00], \( p <0.001 \)), despite the small sample sizes (No Diagnosis = 25 and dementia + Cognitive Disorder, NOS = 21). A score of 24 or lower yielded a sensitivity of 91% and a specificity of 80%. Thus the current study demonstrated possible utility of the SLUMS for discriminating between cognitive normal older adults and those with cognitive impairments.

In sum, the SLUMS is a brief cognitive screener with poor internal consistency, which requires caution in interpreting the final results. Despite the low internal consistency, it significantly correlated with all but one neuropsychological measure (Trails B), including the DRS-2. In fact, the SLUMS score significantly predicted the number of impaired neuropsychological measures (\( R^2 = .19, F(1, 147) = 33.28, \))
Within the context of a heterogeneous clinical comparison group, the SLUMS appeared to be a good screening estimate of the diagnostic performance of a comprehensive neuropsychological battery. This was true for all ROC curve analyses that were conducted. The SLUMS may be sensitive to factors associated with race/ethnicity, so practitioners should use caution when using this measure with multicultural patient populations. Specifically, formal years of education may not be an adequate measure of the quality of one’s education, so clinicians may wish also to evaluate the patient’s word reading skills (i.e., using the WTAR). Regardless, clinicians should be mindful of potential ethnic/racial effects on the SLUMS that may reduce performance. Particular items that demonstrated disparities in the current sample (with African Americans performing lower than European Americans), included calculations \( t(138) = 2.39, p = .02 \), and differences on the logical memory task approached significance \( t(138) = 1.86, p = .07 \). These results suggest that there is evidence for strengths and weaknesses with regard to use of the SLUMS as a screening measure, but more research is needed to understand its utility better.

### 4.1 Clinical Implications

At this time, the research is just being generated regarding use of the SLUMS in a veteran population, for which it was developed. The sensitivity, specificity, PPV, and NPV may vary by setting due to differences in diagnostic base rates, such that diagnostic indicators may be weaker in a general clinical setting and relatively stronger in specialized (i.e., MCI) clinics in which the base rate of the target disorder is higher.

The SLUMS does do a good job discriminating between MCI and no diagnosis and MDD in the specialty clinical setting. Interestingly, when the participants diagnosed with MDD were removed from the “Other Diagnosis” group analyses, the SLUMS did an excellent job discriminating the “Other Diagnosis” group from those with no diagnosis \( (\text{AUC} = 0.93, \text{SE} = 0.04, p < 0.001, 95\% \text{ CI [0.86, 1.00]}) \), with an optimal cutoff score of 24 or less (sensitivity = 91%, specificity = 81%). This was also true for the neuropsychological battery \( (\text{AUC} = 0.93, \text{SE} = 0.04, p < 0.001, 95\% \text{ CI [0.86, 1.00]}) \), with an optimal cutoff of
three deficits that are one standard deviation below the mean (sensitivity = 91%, specificity = 89%). The SLUMS still performed poorly when discriminating between the “Other Diagnosis” group without MDD and the MCI group (AUC = 0.26, SE = 0.06, \( p < 0.001 \), 95% CI [0.13, 0.38]). No cutoff score would provide adequate sensitivity or specificity. In sum, a cutoff score of 25 was optimal for discriminating between MCI and no diagnosis, whereas a slightly lower cutoff score of 24 is recommended for discriminating between MCI and those with MDD given this study’s sample and characteristics.

The benefits and potential harms of screening for MCI must also be considered. These factors also impact the weight that is given to each diagnostic indicator (i.e., is specificity more important than sensitivity? Is PPV more important than other indicators?). The harms of screening individuals who report cognitive difficulties for MCI include the cost of a comprehensive neuropsychological evaluation, which may be used to confirm or refute the presence of MCI, as well as the risk of increasing worry, sadness, or anxiety in the case of a positive cognitive screening (Boustani, Peterson, Hanson, Harris, & Lohr, 2003). This concern may be increased because, as of yet, there is no approved or effective medications for the treatment of MCI (Farlow, 2009; Lin & Neumann, 2013). However, Doody et al. (2009) demonstrated that donepezil, an acetylcholinesterase inhibitor, may slow the rate of progression to dementia, improve aspects of cognition, and result in subjective reports of benefit. It must be noted, though, that Petersen, Stevens, et al. (2005) demonstrated no difference between controls and individuals treated with donepezil in rate of progression to AD over a three-year period.

Earlier identification of MCI also allow individuals to prepare for the possibility of later decline. For example, early detection allows for psychoeducation; discussion of symptoms and prognosis; health, safety, and financial planning; screening of and management of co-existing conditions, improving medication adherence, and preventing medication interactions; and ways in which one can prepare one’s affairs, such as putting in place advanced directives and making decisions about driving. Additionally, potential behavioral, cognitive, and lifestyle treatment options may exist to help slow the rate of pro-
gression or delay the decline in functional abilities (Huckans, Hutson, Twamley, Jak, Kaye, & Storzbach, 2013). Despite numerous limitations of studies of early interventions in MCI, including methodological weaknesses or disregard for quality of life or neuropsychiatric outcomes, Huckans and colleagues concluded that the benefits of cognitive rehabilitation therapies are encouraging, but inconclusive. Petersen, Stevens, and colleagues (2001) argued that sufficient evidence exists to recommend screening individuals with MCI, because these individuals are at increased risk of developing dementia.

Considering the benefits and harms of early screening, it is argued that the risks associated with false positive errors are acceptable when weighed against the benefits of early identification of MCI. Thus, maximizing the sensitivity of cognitive screening measures is recommended. Positive predictive values may be less of a concern, as the risks of a false positive result are believed to be less harmful than that of a false negative. Thus, it may also be important to maximize NPV in addition to sensitivity, so as to increase the chances of accurately capturing individuals with MCI. The SLUMS demonstrated adequate sensitivity (73-81%) when discriminating MCI from no diagnosis or MDD. Sensitivity for the neuropsychological battery and factors scores varied (64-93%) based on the comparison groups and the method (i.e., factor scores, comprehensive neuropsychological battery, cognitive screening measure). Specifically, the memory factor demonstrated the strongest sensitivity (93%) and NPV (71-78%), suggesting that including memory tasks in cognitive screening measures for MCI will be integral to maximizing sensitivity and NPV. The SLUMS demonstrated better NPV (55%) when discriminating MCI from no diagnosis than MCI vs. MDD (44%), but these results suggest that a negative result on the SLUMS may often fail to identify those with MCI.

In sum, it appears as though the SLUMS is best when discriminating between no diagnosis and dementia, Cognitive Disorder NOS, and MCI; or a MCI and MDD. Patients with no diagnosis and a diagnosis of MDD performed similarly, earning an average score of 26, whereas patients with a clinical diagnosis averaged a score of 22 on the SLUMS. The crucial information for clinicians, when determining
whether more evaluation is needed, is the numerous forms of collateral information, such as that obtained from a clinical interview, past medical records, informant reports, and behavioral observations. Even when looking at the number of deficits from the neuropsychological evaluations—which were used to make the original diagnosis—the diagnostic accuracy was comparable to that of the SLUMS. Thus, the validity measures of the SLUMS look promising, but the poor reliability results suggests the need to be cautious in its use as more evaluation of its psychometric and measurement characteristics are needed.
5 REFERENCES


converged. *Archives of Neurology, 66*(3), 300-305.


Kazis, L.E., Miller, D.R., Clark, J., Skinner, K., Lee, A., Rogers, W., ... Linzer, M. (1998). Health-related quality of life in patients served by the Department of Veterans Affairs: Results from the Veterans Health Study. Archives of Internal Medicine, 158, 626-632.


Petersen, R.C., Doody, R., Kurz, A., Mohs, R.C., Morris, J.C., Rabins, P.V., ... Winblad, B. (2001). Current


Tariq, S.H., Tumosa, N., Chibnall, J.T., Perry, M.H., & Morley, J.E. (2006). Comparison of the Saint Louis University Mental Status Examination and the Mini-Mental State Examination for detecting


Zihl, J., Repermund, S., Thum, S., & Unger, K. (2010). Neuropsychological profiles in MCI and in
## VAMC SLUMS Examination

Questions about this assessment tool? E-mail aging@slu.edu

<table>
<thead>
<tr>
<th>Name _______________________________</th>
<th>Age _______________________________</th>
<th>Is patient alert? _______________________________</th>
<th>Level of education _______________________________</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. What day of the week is it?  
2. What is the year?  
3. What state are we in?  
4. Please remember these five objects. I will ask you what they are later.  
   Apple  
   Pen  
   Tie  
   House  
   Car  
5. You have $100 and you go to the store and buy a dozen apples for $3 and a tricycle for $20.  
   (a) How much did you spend?  
   (b) How much do you have left?  
6. Please name as many animals as you can in one minute.  
   (a) 0-4 animals  
   (b) 5-9 animals  
   (c) 10-14 animals  
   (d) 15+ animals  
7. What were the five objects I asked you to remember? I point for each one correct.  
   (a) 87  
   (b) 649  
   (c) 8537  
8. I am going to give you a series of numbers and I would like you to give them to me backwards.  
   For example, if I say 42, you would say 24.  
   (a) 87  
   (b) 649  
   (c) 8537  
9. This is a clock face. Please put in the hour markers and the time at ten minutes to eleven o'clock.  
   (a) Hour markers okay  
   (b) Time correct  
10. Please place an X in the triangle.  
   (a) Which of the above figures is largest?  

11. I am going to tell you a story. Please listen carefully because afterwards, I'm going to ask you some questions about it.  
   Jill was a very successful stockbroker. She made a lot of money on the stock market. She then met Jack, a devastatingly handsome man. She married him and had three children. They lived in Chicago.  
   They then stopped work and stayed at home to bring up her children. When they were teenagers, she went back to work. She and Jack lived happily ever after.  
   (a) What was the female's name?  
   (b) What work did she do?  
   (c) When did she go back to work?  
   (d) What state did she live in?  

---

### TOTAL SCORE

---

### SCORING

<table>
<thead>
<tr>
<th>High School Education</th>
<th>Less Than High School Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>27-30</td>
<td>Normal</td>
</tr>
<tr>
<td>21-26</td>
<td>MNDP</td>
</tr>
<tr>
<td>1-20</td>
<td>Dementia</td>
</tr>
</tbody>
</table>

* Mild Neurocognitive Disorder

APPENDIX B

ROC Curve Analysis for the Factor Scores

**Executive Functions Factor, MCI vs. No Diagnosis.** The AUC for discriminating between those diagnosed with MCI and no diagnosis using tests from the Executive Functions (EF) factor (Digit Span, Trails B, and Digit Symbol-Coding) was .72 with $SE = 0.06$ and 95% CI [.61, .83] ($p = .001$). One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 64%, specificity of 76%, PPV of 89% and NPV of 41%.

**EF Factor, MCI vs. Other Diagnosis.** The AUC for discriminating between those diagnosed with MCI versus another diagnosis using tests from the EF factor was .55 ($SE = 0.06$, $p = .34$, 95% CI [.44, .66]. One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 64%, specificity of 53%, PPV of 69%, and a NPV of 48%.

**EF Factor, MCI vs. MDD.** The AUC for discriminating between those diagnosed with MCI versus MDD using tests from the EF factor was .62 ($SE = 0.07$, $p = .09$, 95% CI [.48, .76]. One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 64%, specificity of 65%, PPV of 86%, and a NPV of 36%.

**Language Factor, MCI vs. No Diagnosis.** The AUC for discriminating between those diagnosed with MCI and no diagnosis using tests from the Language factor (Boston Naming Test, FAS, and WAIS-IV Similarities) was .70 with $SE = 0.06$ and 95% CI [.58, .81] ($p = .003$). One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 65%, specificity of 72%, PPV of 88% and NPV of 41%.

**Language Factor, MCI vs. Other Diagnosis.** The AUC for discriminating between those diagnosed with MCI versus another diagnosis using tests from the Language factor was .57 ($SE = 0.06$, $p = .22$, 95% CI [.46, .68]. One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 65%, specificity of 55%, PPV of 70%, and a NPV of 50%.
**Language Factor, MCI vs. MDD.** The AUC for discriminating between those diagnosed with MCI versus MDD using tests from the Language factor was .76 (SE = 0.05, \( p < .001 \), 95% CI [.65, .86]). One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 65%, specificity of 83%, PPV of 92%, and a NPV of 42%.

**Memory Factor, MCI vs. No Diagnosis.** The AUC for discriminating between those diagnosed with MCI and no diagnosis using tests from the Memory factor (CVLT-II and BVMT-R Delayed Recall scores) was .88 with \( SE = 0.04 \) and 95% CI [.82, .96] (\( p < .001 \)). One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 93%, specificity of 72%, PPV of 91% and NPV of 78%.

**Memory Factor, MCI vs. Other Diagnosis.** The AUC for discriminating between those diagnosed with MCI versus another diagnosis using tests from the EF factor was .60 (SE = 0.06, \( p = .07 \), 95% CI [.49, .71]). One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 93%, specificity of 36%, PPV of 70%, and a NPV of 77%.

**Memory Factor, MCI vs. MDD.** The AUC for discriminating between those diagnosed with MCI versus MDD using tests from the EF factor was .75 (SE = 0.07, \( p < .001 \), 95% CI [.63, .88]). One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 93%, specificity of 52%, PPV of 86%, and a NPV of 71%.

**3-Factor ROC Curve, MCI vs. No Diagnosis.** The AUC for discriminating between those diagnosed with MCI and no diagnosis based on the number of factor composite score deficits was .84 with \( SE = 0.04 \) and 95% CI [.76, .92] (\( p < .001 \)). One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 75%, specificity of 88%, PPV of 95% and NPV of 54%. Based on the results of the ROC curve analyses using the number of impaired measures that comprise the individual factors to discriminate between MCI and no diagnosis, one can infer that the Memory factor may be the factor of interest.
3-Factor ROC Curve, MCI vs. Other Diagnosis. The AUC for discriminating between those diagnosed with MCI versus another diagnosis was .59 ($SE = 0.06, p = .10, 95\% CI [.48, .70]$. One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 75%, specificity of 46%, PPV of 68%, and a NPV of 53%.

3-Factor ROC Curve, MCI vs. MDD. The AUC for discriminating between those diagnosed with MCI versus MDD was .78 ($SE = 0.05, p < .001, 95\% CI [.67,.88]. One deficit resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, one deficit resulted in a sensitivity of 75%, specificity of 74%, PPV of 90%, and NPV of 47%.

ROC Curve Analysis for the DRS-2. Results of the ROC curve analyses on the DRS-2 is only reported for discriminating between MCI and Other Diagnoses, as the other groups (No Diagnosis and MDD) did not have a large enough sample size (three and eight, respectively). The “Other Diagnosis” group could also not be broken down further into more specific diagnostic groups, as the sample sizes for those that completed a DRS-2 were too small (Cognitive Disorder NOS = 3, dementia = 7, Anxiety = 1, MDD = 8).

MCI vs. Other Diagnosis. The AUC for discriminating between those diagnosed with MCI versus another diagnosis was .44 ($SE = 0.09, p = .47, 95\% CI [.26, .61]). A total scaled score of seven or less on the DRS-2 resulted in the most optimal diagnostic indicators for identifying MCI. Specifically, a total scaled score of seven or less resulted in a sensitivity of 30%, specificity of 63%, PPV of 53%, and a NPV of 39%. Overall, the DRS-2 is not a powerful test for discriminating between MCI and other disorders. Most participants with MCI performed in the average or above average range on the DRS-2 (19 out of 27). It is likely that the specificity would be stronger if those diagnosed with a dementia were not included in this “Other Diagnosis” group, as they will likely perform poorly on this brief cognitive measure that was designed to detect dementia. Using the Memory subscale score mildly improved diagnostic indicators, such that sensitivity = 37%, specificity = 63%, PPV = 59%, and NPV = 71%.
Table 1 DSM-IV-TR and NINCDS-ADRDA Criteria for Diagnosing Dementia and Alzheimer’s Disease

<table>
<thead>
<tr>
<th>Dementia</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR)a | -Multiple cognitive deficits that include memory impairment and at least one of the following cognitive disturbances: aphasia, apraxia, agnosia, or a disturbance in executive functioning  
-Deficits cause significant impairment in occupational or social functioning  
-Must represent a decline from a previously higher level of functioning  
-Do not diagnose with dementia if cognitive deficits occur exclusively during the course of a delirium  
-Dementia may be etiologically related to a general medical condition, to the persisting effects of a substance use (including toxin exposure), or to a combination of these factors |

| Alzheimer’s Disease                                                      | Clinical diagnosis of PROBABLE Alzheimer’s disease:  
-Dementia established by clinical and neuropsychological examination, which includes deficits in two or more areas of cognition.  
-Progressive worsening of memory and other cognitive deficits  
-Onset between the ages of 40 and 90  
-Absence of other diseases that could account for the deficits in memory and cognition |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA)b</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Sample Demographics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Percent of Sample (N = 148)</th>
<th>Average SLUMS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean [SD])</td>
<td>68.48 (7.64)</td>
<td></td>
</tr>
<tr>
<td>Education (mean [SD])</td>
<td>13.24 (2.82)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European American</td>
<td>68</td>
<td>23.60 (4.10)</td>
</tr>
<tr>
<td>African American</td>
<td>32</td>
<td>21.64 (4.63)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI</td>
<td>50.7</td>
<td>22.43 (3.32)</td>
</tr>
<tr>
<td>Major Depression Disorder</td>
<td>15.5</td>
<td>26.0 (3.50)</td>
</tr>
<tr>
<td>Dementia</td>
<td>8.8</td>
<td>18.50 (4.93)</td>
</tr>
<tr>
<td>Cognitive Disorder, NOS</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>No Diagnosis</td>
<td>16.9</td>
<td>26.24 (2.70)</td>
</tr>
<tr>
<td>Vascular Comorbidities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>78</td>
<td>23.01 (4.24)</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>62</td>
<td>22.89 (4.25)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>34</td>
<td>22.86 (4.74)</td>
</tr>
</tbody>
</table>

Note. SLUMS = Saint Louis University Mental Status.
Table 3 Correlations Between SLUMS Score and Demographic Factors and Neuropsychological Test Scores

<table>
<thead>
<tr>
<th></th>
<th>SLUMS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td>-.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Education</td>
<td>.11</td>
<td>.16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. DRS-2 Total Score</td>
<td>.44**</td>
<td>-.14</td>
<td>.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. WTAR FSIQ</td>
<td>.30**</td>
<td>.24**</td>
<td>.62**</td>
<td>.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. WTAR VIQ</td>
<td>.29**</td>
<td>.23**</td>
<td>.63**</td>
<td>.11</td>
<td>1**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. WTAR PIQ</td>
<td>.31**</td>
<td>.22**</td>
<td>.63**</td>
<td>.10</td>
<td>.99**</td>
<td>.99**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. CVLT LD Free Recall</td>
<td>.46**</td>
<td>-.13</td>
<td>.18**</td>
<td>.51**</td>
<td>.22*</td>
<td>.23**</td>
<td>.22**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Digit Span</td>
<td>.45**</td>
<td>-.02</td>
<td>.25*</td>
<td>N/A</td>
<td>.48**</td>
<td>.49**</td>
<td>.49**</td>
<td>.49**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Digit Symbol-Coding</td>
<td>.34**</td>
<td>.42**</td>
<td>.32**</td>
<td>N/A</td>
<td>.40**</td>
<td>.40**</td>
<td>.41**</td>
<td>.18</td>
<td>.39**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10. Similarities</td>
<td>.42**</td>
<td>.21</td>
<td>.46**</td>
<td>N/A</td>
<td>.63**</td>
<td>.63**</td>
<td>.64**</td>
<td>.28*</td>
<td>.44**</td>
<td>.45**</td>
<td>1</td>
</tr>
<tr>
<td>11. Boston Naming Test</td>
<td>.34**</td>
<td>0</td>
<td>-.09</td>
<td>.02</td>
<td>.28**</td>
<td>.29**</td>
<td>.27**</td>
<td>.31**</td>
<td>.37**</td>
<td>.28**</td>
<td>.36**</td>
</tr>
<tr>
<td>12. Phonemic Fluency (FAS)</td>
<td>.32**</td>
<td>.01</td>
<td>.04</td>
<td>.49**</td>
<td>.23**</td>
<td>.25**</td>
<td>.24**</td>
<td>.38**</td>
<td>.49**</td>
<td>.40**</td>
<td>.35**</td>
</tr>
<tr>
<td>13. Animal Fluency</td>
<td>.39**</td>
<td>-.13</td>
<td>-.03</td>
<td>.50**</td>
<td>0</td>
<td>.02</td>
<td>0</td>
<td>.42**</td>
<td>.27**</td>
<td>.29**</td>
<td>.13</td>
</tr>
<tr>
<td>14. BVMT-R Delayed Recall</td>
<td>.40**</td>
<td>-.09</td>
<td>.20*</td>
<td>.35*</td>
<td>.36**</td>
<td>.36**</td>
<td>.36**</td>
<td>.43**</td>
<td>.41**</td>
<td>.32**</td>
<td>.23**</td>
</tr>
<tr>
<td>15. Trails B</td>
<td>.14</td>
<td>-.08</td>
<td>-.22*</td>
<td>.61**</td>
<td>-.07</td>
<td>-.06</td>
<td>-.07</td>
<td>.24**</td>
<td>.41**</td>
<td>.27*</td>
<td>.06</td>
</tr>
<tr>
<td>16. Trails A</td>
<td>.25**</td>
<td>-.04</td>
<td>-.05</td>
<td>.44**</td>
<td>.06</td>
<td>.07</td>
<td>.05</td>
<td>.24**</td>
<td>.34**</td>
<td>.48**</td>
<td>.13</td>
</tr>
<tr>
<td>17. ACBS Score</td>
<td>.11</td>
<td>-.15</td>
<td>-.05</td>
<td>.10</td>
<td>-.05</td>
<td>-.05</td>
<td>-.04</td>
<td>.08</td>
<td>-.09</td>
<td>-.03</td>
<td>-.11</td>
</tr>
<tr>
<td>18. Medication Count</td>
<td>.05</td>
<td>-.16</td>
<td>0</td>
<td>.05</td>
<td>-.02</td>
<td>-.02</td>
<td>-.01</td>
<td>.18*</td>
<td>-.03</td>
<td>.05</td>
<td>-.02</td>
</tr>
<tr>
<td>19. BDI-2</td>
<td>.32*</td>
<td>-.39**</td>
<td>-.33*</td>
<td>-.01</td>
<td>-.07</td>
<td>-.07</td>
<td>-.08</td>
<td>.14</td>
<td>0</td>
<td>.06</td>
<td>.10</td>
</tr>
<tr>
<td>20. # Neuropsych Deficits</td>
<td>-.43**</td>
<td>-.06</td>
<td>-.06</td>
<td>-.54**</td>
<td>-.28**</td>
<td>-.28**</td>
<td>-.28**</td>
<td>-.58**</td>
<td>-.67**</td>
<td>-.63**</td>
<td>-.44**</td>
</tr>
<tr>
<td>21. EF Factor</td>
<td>.32**</td>
<td>.01</td>
<td>.03</td>
<td>.61**</td>
<td>.22**</td>
<td>.23**</td>
<td>.23**</td>
<td>.38**</td>
<td>.78**</td>
<td>.74**</td>
<td>.45**</td>
</tr>
<tr>
<td>22. Language Factor</td>
<td>.44**</td>
<td>.05</td>
<td>.03</td>
<td>.30*</td>
<td>.38**</td>
<td>.39**</td>
<td>.38**</td>
<td>.44**</td>
<td>.56**</td>
<td>.48**</td>
<td>.71**</td>
</tr>
<tr>
<td>23. Memory Factor</td>
<td>.50**</td>
<td>-.13</td>
<td>.20*</td>
<td>-.53**</td>
<td>.32**</td>
<td>.32**</td>
<td>.32**</td>
<td>.82**</td>
<td>.51**</td>
<td>.28**</td>
<td>.22**</td>
</tr>
</tbody>
</table>

** p < 0.01, * p < 0.05. SLUMS = Saint Louis University Mental Status; DRS = Dementia Rating Scale; WTAR = Wechsler Test of Adult Reading; FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; LD = Long Delay; BVMT-R = Brief Visual Memory Test-Revised; GDS = Geriatric Depression Scale; Neuropsych = Neuropsychological Battery
Table 3 Correlations (Continued)

<table>
<thead>
<tr>
<th></th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. DRS-2 Total Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. WTAR FSIQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. WTAR VIQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. WTAR PIQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. CVLT Long Delay Free Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Digit Span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Digit Symbol-Coding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Similarities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Boston Naming Test</td>
<td>.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Phonemic Fluency (FAS)</td>
<td>.37**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Animal Fluency</td>
<td>.42**</td>
<td>.47**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. BVMT-R Delayed Recall</td>
<td>.32**</td>
<td>.25**</td>
<td>.24**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Trails B</td>
<td>.23**</td>
<td>.37**</td>
<td>.24**</td>
<td>.24**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Trails A</td>
<td>.26**</td>
<td>.38**</td>
<td>.43**</td>
<td>.24**</td>
<td>.58**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. ACBS Score</td>
<td>-.02</td>
<td>-.07</td>
<td>-.04</td>
<td>.06</td>
<td>-.01</td>
<td>-.01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Medication Count</td>
<td>.11</td>
<td>-.10</td>
<td>-.03</td>
<td>.24**</td>
<td>-.02</td>
<td>0</td>
<td>.64**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. BDI-2</td>
<td>.27</td>
<td>.26</td>
<td>.23</td>
<td>.22</td>
<td>.33*</td>
<td>.24</td>
<td>.45**</td>
<td>.42**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>20. # Neuropsyc Deficits</td>
<td>-.52**</td>
<td>-.61**</td>
<td>-.56**</td>
<td>-.55**</td>
<td>-.60**</td>
<td>-.57**</td>
<td>-.04</td>
<td>-.04</td>
<td>-.38**</td>
<td>1</td>
</tr>
<tr>
<td>21. EF Factor</td>
<td>-.31**</td>
<td>.50**</td>
<td>.27**</td>
<td>.40**</td>
<td>.87**</td>
<td>.59**</td>
<td>.02</td>
<td>.03</td>
<td>.34</td>
<td>-.75**</td>
</tr>
<tr>
<td>22. Language Factor</td>
<td>.81**</td>
<td>.78**</td>
<td>.50**</td>
<td>.34**</td>
<td>.36**</td>
<td>.37**</td>
<td>-.07</td>
<td>.01</td>
<td>.29*</td>
<td>-.67**</td>
</tr>
<tr>
<td>23. Memory Factor</td>
<td>.37**</td>
<td>.36**</td>
<td>.41**</td>
<td>.85**</td>
<td>.29**</td>
<td>.28**</td>
<td>.11</td>
<td>.27**</td>
<td>.22</td>
<td>-.66**</td>
</tr>
</tbody>
</table>

Note. Correlations among factors: EF and Language (.52**) and Memory (.45**), Language and Memory (.43**).
Table 4 Inter-Item Correlations for the SLUMS

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Day</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Year</td>
<td>-.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Math</td>
<td>.09</td>
<td>.17</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Fluency</td>
<td>.07</td>
<td>.05</td>
<td>.21</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Digit Span</td>
<td>.06</td>
<td>.27</td>
<td>.33</td>
<td>.25</td>
<td>.31</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Clock</td>
<td>-.06</td>
<td>-.01</td>
<td>.03</td>
<td>.25</td>
<td>.09</td>
<td>.30</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Triangle</td>
<td>.27</td>
<td>-.04</td>
<td>.15</td>
<td>.13</td>
<td>.15</td>
<td>.21</td>
<td>-.04</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11. Logical Memory</td>
<td>.05</td>
<td>.12</td>
<td>.18</td>
<td>.37</td>
<td>.26</td>
<td>.25</td>
<td>.21</td>
<td>.09</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. Item 3 was removed due to zero variance. Item 4 is not scored on the SLUMS. SLUMS = Saint Louis University Mental Status.
Table 5 Item-Total Statistics for the SLUMS

<table>
<thead>
<tr>
<th></th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach’s Alpha if Item Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Day</td>
<td>.11</td>
<td>.09</td>
<td>.58</td>
</tr>
<tr>
<td>2. Year</td>
<td>.19</td>
<td>.11</td>
<td>.57</td>
</tr>
<tr>
<td>5. Math</td>
<td>.30</td>
<td>.16</td>
<td>.53</td>
</tr>
<tr>
<td>6. Fluency</td>
<td>.43</td>
<td>.21</td>
<td>.51</td>
</tr>
<tr>
<td>7. Recall</td>
<td>.36</td>
<td>.17</td>
<td>.51</td>
</tr>
<tr>
<td>8. Digit Span</td>
<td>.48</td>
<td>.31</td>
<td>.50</td>
</tr>
<tr>
<td>9. Clock</td>
<td>.24</td>
<td>.17</td>
<td>.55</td>
</tr>
<tr>
<td>10. Triangle</td>
<td>.18</td>
<td>.14</td>
<td>.57</td>
</tr>
<tr>
<td>11. Logical Memory</td>
<td>.40</td>
<td>.20</td>
<td>.52</td>
</tr>
</tbody>
</table>

Note. SLUMS = Saint Louis University Mental Status.
Table 6 Mean Sample Score on Each Neuropsychological Measure

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean* (Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAIS-IV VIQ</td>
<td>97.76 (12.83)</td>
</tr>
<tr>
<td>WAIS-IV PIQ</td>
<td>96.26 (10.58)</td>
</tr>
<tr>
<td>WAIS-IV FSIQ</td>
<td>96.77 (13.02)</td>
</tr>
<tr>
<td>WAIS-IV Digit Span</td>
<td>-0.63 (0.83)</td>
</tr>
<tr>
<td>WAIS-IV Digit Symbol-Coding</td>
<td>-0.75 (0.82)</td>
</tr>
<tr>
<td>WAIS-IV Similarities</td>
<td>-0.18 (0.90)</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>-0.15 (1.15)</td>
</tr>
<tr>
<td>Phonemic Fluency (FAS)</td>
<td>-0.67 (1.05)</td>
</tr>
<tr>
<td>Animal Fluency</td>
<td>-0.39 (1.18)</td>
</tr>
<tr>
<td>BVMT-R Delayed Recall</td>
<td>-1.10 (1.23)</td>
</tr>
<tr>
<td>CVLT-II Long Delay Free Recall</td>
<td>-0.72 (1.14)</td>
</tr>
<tr>
<td>Trails A</td>
<td>-0.64 (1.07)</td>
</tr>
<tr>
<td>Trails B</td>
<td>-0.37 (1.07)</td>
</tr>
<tr>
<td>DRS-2 Total Score</td>
<td>8.61 (2.53)</td>
</tr>
</tbody>
</table>

*All values are z-scores, with the exception of Verbal IQ (VIQ), Performance IQ (PIQ), and Full Scale IQ (FSIQ; standard scores) and the Dementia Rating Scale, 2nd Edition (DRS-2) Total Score (scaled score). SLUMS = Saint Louis University Mental Status; BVMT-R = Brief Visuospatial Memory Test, Revised; CVLT-II = California Verbal Learning Test, Second Edition.
Table 7 Cutoff Scores for Mild Cognitive Impairment on the SLUMS

<table>
<thead>
<tr>
<th>Cutoff Score (≤)</th>
<th>MCI vs. No Diagnosis</th>
<th>MCI vs. Other Diagnosis</th>
<th>MCI vs. Major Depressive Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>1-Specificity</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>31.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>29.5</td>
<td>1.00</td>
<td>0.92</td>
<td>1.00</td>
</tr>
<tr>
<td>28.5</td>
<td>0.99</td>
<td>0.80</td>
<td>0.99</td>
</tr>
<tr>
<td>27.5</td>
<td>0.95</td>
<td>0.64</td>
<td>0.95</td>
</tr>
<tr>
<td>26.5</td>
<td>0.88</td>
<td>0.48</td>
<td>0.88</td>
</tr>
<tr>
<td>25.5</td>
<td>0.81</td>
<td>0.32</td>
<td>0.81</td>
</tr>
<tr>
<td>24.5</td>
<td>0.73</td>
<td>0.20</td>
<td>0.73</td>
</tr>
<tr>
<td>23.5</td>
<td>0.59</td>
<td>0.16</td>
<td>0.59</td>
</tr>
<tr>
<td>22.5</td>
<td>0.48</td>
<td>0.12</td>
<td>0.48</td>
</tr>
<tr>
<td>21.5</td>
<td>0.36</td>
<td>0.04</td>
<td>0.36</td>
</tr>
<tr>
<td>20.5</td>
<td>0.27</td>
<td>0.04</td>
<td>0.27</td>
</tr>
<tr>
<td>19.5</td>
<td>0.19</td>
<td>0.04</td>
<td>0.19</td>
</tr>
<tr>
<td>18.5</td>
<td>0.19</td>
<td>0</td>
<td>0.11</td>
</tr>
<tr>
<td>17.5</td>
<td>0.11</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>16.5</td>
<td>0.04</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>15.5</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>14.5</td>
<td>--</td>
<td>--</td>
<td>0.00</td>
</tr>
<tr>
<td>12.5</td>
<td>--</td>
<td>--</td>
<td>0.00</td>
</tr>
<tr>
<td>10.5</td>
<td>--</td>
<td>--</td>
<td>0.00</td>
</tr>
<tr>
<td>9.0</td>
<td>--</td>
<td>--</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. MCI = Mild Cognitive Impairment; SLUMS = Saint Louis University Mental Status.
Table 8 Cutoff Scores (Number of Deficits) for MCI using a Neuropsychological Battery

<table>
<thead>
<tr>
<th>Cutoff Score (≥)</th>
<th>MCI vs. No Diagnosis</th>
<th>MCI vs. Other Diagnosis</th>
<th>MCI vs. Major Depressive Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>1-Specificity</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>-1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.50</td>
<td>1.00</td>
<td>0.64</td>
<td>1.00</td>
</tr>
<tr>
<td>1.50</td>
<td>0.91</td>
<td>0.44</td>
<td>0.91</td>
</tr>
<tr>
<td>2.50</td>
<td>0.79</td>
<td>0.12</td>
<td>0.79</td>
</tr>
<tr>
<td>3.50</td>
<td>0.56</td>
<td>0.04</td>
<td>0.56</td>
</tr>
<tr>
<td>4.50</td>
<td>0.31</td>
<td>0.04</td>
<td>0.31</td>
</tr>
<tr>
<td>5.50</td>
<td>0.17</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>6.50</td>
<td>0.08</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>7.50</td>
<td>0.04</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>8.50</td>
<td>0.03</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>10.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. MCI = Mild Cognitive Impairment; SLUMS = Saint Louis University Mental Status.
Table 9 Cutoff Scores (Number of Deficits for MCI Using Factor Scores)

<table>
<thead>
<tr>
<th>Cutoff Score (≥)</th>
<th>MCI vs. No Diagnosis</th>
<th>MCI vs. Other Diagnosis</th>
<th>MCI vs. Major Depressive Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>1-Specificity</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>-1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.50</td>
<td>0.75</td>
<td>0.12</td>
<td>0.75</td>
</tr>
<tr>
<td>1.50</td>
<td>0.36</td>
<td>0.00</td>
<td>0.36</td>
</tr>
<tr>
<td>2.50</td>
<td>0.08</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>3.50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. MCI = Mild Cognitive Impairment; SLUMS = Saint Louis University Mental Status.
Table 10 Diagnostic Indicators for the SLUMS, the Comprehensive Neuropsychological Battery, the Factor Scores, and the DRS-2

<table>
<thead>
<tr>
<th></th>
<th>Cutoff Score</th>
<th>AUC (SE)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SLUMS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI vs. No Dx</td>
<td>25</td>
<td>.82 (0.05)</td>
<td>81</td>
<td>68</td>
<td>88</td>
<td>55</td>
<td>1.26</td>
</tr>
<tr>
<td>MCI vs. Other Dx</td>
<td>24</td>
<td>.51 (0.06)</td>
<td>73</td>
<td>40</td>
<td>73</td>
<td>49</td>
<td>0.06</td>
</tr>
<tr>
<td>MCI vs. MDD</td>
<td>24</td>
<td>.79 (0.06)</td>
<td>73</td>
<td>70</td>
<td>89</td>
<td>44</td>
<td>1.04</td>
</tr>
<tr>
<td><strong>Neuropsychological Battery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI vs. No Dx</td>
<td>3</td>
<td>.89 (0.04)</td>
<td>79</td>
<td>88</td>
<td>95</td>
<td>59</td>
<td>1.65</td>
</tr>
<tr>
<td>MCI vs. Other Dx</td>
<td>4</td>
<td>.59 (0.06)</td>
<td>56</td>
<td>56</td>
<td>67</td>
<td>44</td>
<td>0.26</td>
</tr>
<tr>
<td>MCI vs. MDD</td>
<td>3</td>
<td>.79 (0.06)</td>
<td>79</td>
<td>71</td>
<td>90</td>
<td>48</td>
<td>1.14</td>
</tr>
<tr>
<td><strong>EF Factor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI vs. No Dx</td>
<td>1</td>
<td>.72 (0.06)</td>
<td>64</td>
<td>76</td>
<td>89</td>
<td>41</td>
<td>1.26</td>
</tr>
<tr>
<td>MCI vs. Other Dx</td>
<td>1</td>
<td>.55 (0.06)</td>
<td>64</td>
<td>53</td>
<td>69</td>
<td>48</td>
<td>0.15</td>
</tr>
<tr>
<td>MCI vs. MDD</td>
<td>1</td>
<td>.62 (0.07)</td>
<td>64</td>
<td>65</td>
<td>86</td>
<td>36</td>
<td>0.67</td>
</tr>
<tr>
<td><strong>Language Factor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI vs. No Dx</td>
<td>1</td>
<td>.70 (0.06)</td>
<td>65</td>
<td>72</td>
<td>88</td>
<td>41</td>
<td>1.13</td>
</tr>
<tr>
<td>MCI vs. Other Dx</td>
<td>1</td>
<td>.57 (0.06)</td>
<td>65</td>
<td>55</td>
<td>70</td>
<td>50</td>
<td>0.10</td>
</tr>
<tr>
<td>MCI vs. MDD</td>
<td>1</td>
<td>.76 (0.05)</td>
<td>65</td>
<td>83</td>
<td>92</td>
<td>42</td>
<td>1.07</td>
</tr>
<tr>
<td><strong>Memory Factor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI vs. No Dx</td>
<td>1</td>
<td>.88 (0.04)</td>
<td>93</td>
<td>72</td>
<td>91</td>
<td>78</td>
<td>1.54</td>
</tr>
<tr>
<td>MCI vs. Other Dx</td>
<td>1</td>
<td>.60 (0.06)</td>
<td>93</td>
<td>36</td>
<td>70</td>
<td>77</td>
<td>0.40</td>
</tr>
<tr>
<td>MCI vs. MDD</td>
<td>1</td>
<td>.75 (0.07)</td>
<td>93</td>
<td>52</td>
<td>86</td>
<td>71</td>
<td>1.08</td>
</tr>
<tr>
<td><strong>Factor Composite</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI vs. No Dx</td>
<td>1</td>
<td>.84 (0.04)</td>
<td>75</td>
<td>88</td>
<td>95</td>
<td>54</td>
<td>1.56</td>
</tr>
<tr>
<td>MCI vs. Other Dx</td>
<td>1</td>
<td>.59 (0.06)</td>
<td>75</td>
<td>46</td>
<td>68</td>
<td>53</td>
<td>0.25</td>
</tr>
<tr>
<td>MCI vs. MDD</td>
<td>1</td>
<td>.78 (0.05)</td>
<td>75</td>
<td>74</td>
<td>90</td>
<td>47</td>
<td>1.17</td>
</tr>
<tr>
<td><strong>DRS-2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCI vs. Other Dx</td>
<td>7</td>
<td>.44 (0.09)</td>
<td>30</td>
<td>63</td>
<td>53</td>
<td>39</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Note. SLUMS = Saint Louis University Mental Status; DRS-2 = Dementia Rating Scale, Second Edition; PPV = Positive Predictive Value; NPV = Negative Predictive Value; MCI = Mild Cognitive Impairment; Dx = Diagnosis; EF = Executive Functions
Figure 1 Initial Confirmatory Factor Analysis Model

Figure 2 Final Confirmatory Factor Analysis Model