Shwartz, Morris, and Penna, 2019. Psychometric properties of the Saint Louis University Mental Status Examination.pdf

Susan Shwartz, Shepherd Center

Available at: https://works.bepress.com/susan-shwartz/1/
Psychometric properties of the Saint Louis University Mental Status Examination

Susan K. Shwartz, Robin D. Morris & Suzanne Penna

To cite this article: Susan K. Shwartz, Robin D. Morris & Suzanne Penna (2019) Psychometric properties of the Saint Louis University Mental Status Examination, Applied Neuropsychology: Adult, 26:2, 101-110, DOI: 10.1080/23279095.2017.1362407

To link to this article: https://doi.org/10.1080/23279095.2017.1362407

Published online: 19 Sep 2017.
Psychometric properties of the Saint Louis University Mental Status Examination

Susan K. Shwartz*a,b,c, Robin D. Morrisb and Suzanne Penna*b,d

*aDepartment of Neurology, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia, USA; bDepartment of Psychology, Georgia State University, Atlanta, Georgia, USA; cAcquired Brain Injury Program, Shepherd Center - Shepherd Pathways Division, Decatur, Georgia, USA; dDepartment of Rehabilitation Medicine, Emory University School of Medicine, Atlanta, GA, USA

ABSTRACT

The Saint Louis University Mental Status (SLUMS) Examination is a brief screening measure for mild neurocognitive disorder developed for use with veterans. To date, there has been a paucity of research on its psychometric properties. The purpose of this study is to evaluate the psychometric properties of the SLUMS in a referred sample to a specialty clinic. Using a sample of 148 male veterans referred to a Mild Cognitive Impairment (MCI) Clinic for evaluation, the ability of the SLUMS to discriminate between MCI versus Major Depressive Disorder (MDD) or no diagnosis was compared to results from a more comprehensive neuropsychological battery. The SLUMS was significantly correlated with every neuropsychological measure (r = .25 to .46), except for Trails B (r = .14). Diagnostic discriminability was comparable to a typical, longer, neuropsychological battery for discriminating between MCI and no diagnosis, and between MCI and MDD. The study provides additional psychometric support for the SLUMS as a viable brief cognitive screening measure in veteran populations, particularly when discriminating between MCI and MDD.

KEYWORDS

Cognitive screening measure; discriminability; mild cognitive impairment; psychometrics; veterans

Intervention research in the area of neurodegenerative disorders has begun to focus on identifying patients at 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 is thought to provide the best opportunity to intervene in the disease process and potentially delay or reverse a neurodegenerative process, particularly if treatable etiologies are identified. However, there is still significant difficulty in the ability to identify accurately and efficiently the prodromal stages of a dementing process in general practice settings (e.g., Mitchell, 2009).

In the absence of costly and invasive biomarkers typically reserved for research studies, the best way to accurately identify an early neurodegenerative disorder is through a comprehensive neuropsychological assessment (Sperling et al., 2011). Neuropsychological testing is a well-validated tool to accurately identify a neurodegenerative disorder, even more so if serial assessments are completed. Such assessments can also be used to rule out other possible explanations for cognitive changes. However, such comprehensive assessments are very time consuming and costly, especially given the low prevalence of such disorders in the general population. Thus, more time-efficient, sensitive, and specific screening measures are needed to identify those most at risk for cognitive impairment. Individuals who are identified by such screening measures can then be referred for more extensive assessment.

The Saint Louis University Mental Status Examination (SLUMS; Tariq, Tumosa, Chibnall, Perry, & Morley, 2006) is a neurocognitive screening measure that purports to be a sensitive measure for detecting early cognitive declines in the veteran population. Veterans who use the Veterans Health Administration (VHA) as their primary source of care constitute a specialized population (Morgan, Teal, Reddy, Ford, & Ashton, 2005; National Center for Veterans Analysis and Statistics, 2011a, 2011b). These individuals are more likely to be older, poorer, unemployed or underemployed, African American, are more likely to report poorer physical and mental health, 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, Lofgren, VanRuiswyk, & Layde, 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). Because of these factors, the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), the...
most commonly used screening measure for MCI, may have a too high cutoff score for a veteran population, which results in too many false positives (Whitney, Mossbarger, Herman, & Ibarra, 2012). Thus, the SLUMS was developed to better capture the cognitive characteristics of the aging veteran population (Tariq et al., 2006). However, there is limited psychometric data to support its use for detecting MCI outside the developer's research lab (Brown, Lawson, McDaniel, & Wildman, 2012; Cruz-Oliver, Malmstrom, Allen, Tumosa, & Morley, 2012; Cummings-Vaughn, Cruz-Oliver, Malmstrom, Tumosa, & Morley, 2012; Hawkins, Kilian, Firek, Kashner, Firek, & Silvet, 2012; Stewart, O’Riley, Edelstein, & Gould, 2012; Tariq et al., 2006).

The Saint Louis University Mental Status Examination

The SLUMS is an 11-item cognitive screener that yields an overall score out of 30 possible points, similar to both the Mini Mental Status Exam (MMSE; Folstein, Folstein, & McHugh, 1975) and MoCA. 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 and colleagues did not report utilizing a comprehensive neuropsychological examination to diagnose Mild Neurocognitive Disorder (MNCD), but instead diagnose MNCD and dementia based on history, a physical examination, and neuropsychological measures with the MMSE using receiver operating characteristic (ROC) curve analyses. For those with at least a high school education, the following cut-scores were reported: 27–30 is Normal, 21–26 suggests MNCD, and 1–20 suggests Dementia. For those with less than a high school education, the following cut-scores were reported: 25–30 is Normal, 20–24 suggests MNCD, and 1–19 suggests dementia. The area under the curve (AUC), a measure of overall diagnostic accuracy, was 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. 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).

The SLUMS has a number of strengths and weaknesses. The SLUMS provides suggested cutoff scores for MNCD and dementia and 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 that the MMSE does not, such as verbal fluency and executive functioning. Furthermore, the SLUMS has a too high cutoff score (<26) for a veteran population, which results in a high number of false positive errors (Whitney et al., 2012). Whitney et al. reported that the specificity among middle-aged veterans was poor for the MoCA (57%). Using a lower cutoff score of ≤23 yielded a great specificity (75%), but decreased sensitivity from 86% to 72%. This modified sensitivity level, though, was still far greater than that for 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 (≤20).

The most important weakness of the SLUMS is the paucity of psychometric data regarding its validity and reliability. As noted above, the majority of studies utilizing the SLUMS are from the same research group that developed the measure. Only one study has validated the SLUMS by demonstrating good correlations with a battery of common neuropsychological measures and another validated cognitive screener, the MMSE (Feliciano et al., 2013). They found that 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. A second study by Wildman and McDaniel (2014) provided a validation sample of the SLUMS with nonveterans seeking disability, and found adequate concurrent validity in that population when comparing SLUMS to memory measures (e.g., Wechsler Memory Scale). While the prior studies demonstrate concurrent validity for the SLUMS with a neuropsychological battery or specifically to measures of learning and memory; they do not examine the validity of the SLUMS in a specialty (MCI) clinic sample, nor did the study 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. Depression in older adults is associated with cognitive impairment, as symptoms of...
depression include difficulty with concentration and impaired memory. The ability to provide differential diagnosis between a possible neurodegenerative cognitive impairment and that of depression has long been an important need in clinical services.

The current study improves upon previous research in a number of ways. First, we validated the SLUMS in an independent veteran sample diagnosed with MCI and provide additional psychometric information regarding the measure, that is, sensitivity, specificity, positive predictive values and negative predictive values. Second, in our study, a comprehensive neuropsychological battery was administered to support the MCI diagnosis, which is considered the “gold standard” for diagnosis of MCI (Cruz-Oliver et al., 2012). Third, the current study evaluates the criterion validity of the measure when differentiating between MCI and no diagnosis, and between MCI and depression; as the latter discrimination has been one of the most difficult challenges for medical and mental health professionals working with older adults (Steffens, 2008).

Method

Procedures

Archival data was used that was continuously collected from the Mild Cognitive Impairment (MCI) Clinic at the Atlanta Veterans Affairs Medical Center (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. A SLUMS was required to be completed prior to referral to the MCI Clinic.

Results of the brief cognitive screener and all available medical histories and previous evaluations were reviewed for all patients prior to their scheduled visits. 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.

Anticholinergic medications are also known to impact cognition negatively in older adults. To document such medications and determine if they significantly impacted performance in this sample, two independent lists were used to calculate a medication-related cognitive effect score (American Geriatrics Society [AGS] Beers Criteria, 2012; 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). 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). Furthermore, grouping the number of anticholinergic medications taken (i.e., 1 = no medications, 2 = one or two medications, 3 = three or more) and conducting Spearman rank order correlations also yielded non-significant results.

In addition to a brief medical evaluation, a neuropsychological assessment was completed. Only participants who passed a stand-alone performance validity measure (TOMM; Tombaugh, 1996) were included in the final analyses. Seven participants were excluded who exhibited suboptimal performance validity.

Participants

The majority of patients seen in the MCI Clinic met the following inclusion criteria: a) ages 50–69, b) subjective memory complaint from patient and/or 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 three months), and g) have 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 concussion, 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.

With regard to neuropsychological testing, participants completed one measure from each cognitive domain (intelligence, attention, language, memory [visual and/or verbal], processing speed, and executive functioning). A subset of the sample additionally completed the Dementia Rating Scale-2 (n = 49). These measures were used by the clinical neuropsychologist
to make a clinical diagnosis along with all other available patient clinical, laboratory and historical information, excluding the SLUMS score. Participants with depression as a primary diagnosis, and not diagnosed with MCI, were included to test the differential diagnostic utility of the SLUMS for MCI and depression.

Demographics of the final sample are summarized in Table 1. The final sample consisted of 148 male (0 female) participants with a mean age of 68.48 (range = 50–88, SD = 7.64). 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 completed at least high school or higher. Those with a high school education and higher did not differ significantly in their performance on the SLUMS (M = 23.18, SD = 4.28) from those with less than a high school education (M = 22.04, SD = 4.68, t(146) = 1.31, p = .23). Education was not correlated with the SLUMS (r = .12, p = .19).

The sample was approximately two-thirds European American (n = 101, 68%) and one 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 relationship 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 (Story Memory) than European American participants, which accounted for the mean differences between the two groups.

**Measures**

The following measures were utilized by the VA clinicians to make a clinical diagnosis and were included in statistical analyses. All scores included in the statistical analyses were converted from the respective normative sample standard scores to standardized z-scores. The SLUMS was not used to make a diagnosis so as to avoid circularity bias.

*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. The WTAR was co-normed with the WAIS-III and provides estimates of WAIS-III Full Scale IQ scores, which were used in analyses to determine discriminant validity.

*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 Total Scaled Score was used in analyses.

*Trail Making Test A* (Reitan & Wolfson, 1985) is a visual-motor task that involves attention, visual scanning, motor speed, and sequencing skills. Examinees are required to draw a line in numerical order from number 1 through number 25 as quickly as possible.

*The Digit Symbol-Coding subtest of the WAIS-IV is a timed transcription task that* was included as a measure of processing speed. The score is based on the number of correct digit-symbol pairings completed in the time limit.

*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) in 60 seconds. Animal Fluency is a similar measure that provides a semantic structure. Performance on these measures is based on the total number of words generated, which yields a final score that was used in statistical analyses. The COWAT is often considered an executive functioning test that assesses verbal fluency (e.g., Alvarez & Emory, 2006).

*The Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 2001)* is a confrontation naming task with 60 line-drawn objects that a person must name. The total number correct is based on the number of spontaneously named items plus the number of items named with a semantic cue.

*California Verbal Learning Test, Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000)* is a measure of verbal learning and memory. A list of 16 words that can be organized into four different categories is recited over five trials. This is followed by short and long (20-minute) free and cued delay trials.

---

**Table 1. Sample demographics.**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Percentage of sample (N = 148)</th>
<th>Average SLUMS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>100% male</td>
<td></td>
</tr>
<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 (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>78</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.*
The Long Delay Free Recall score was included in statistical analyses.

The Brief Visual Memory Test, Revised (BVMT-R; Benedict, 1997) was used to assess visual learning and memory. Six designs are presented for 10 seconds over three trials. Following a 25-minute delay, the examinee recalls as many figures as they can. A recognition task and optional copy trial is included. The Delayed Recall Score was used in analyses.

Similarities subtest of the WAIS-IV is a measure of abstract verbal reasoning. The total subtest score was used in analyses.

Trail Making Test B (Reitan & Wolfson, 1985) is a measure of cognitive set shifting. Examinees are asked to connect numbers and letters in numerical and alphabetical order, switching between numbers and letters as quickly as possible.

The Beck Depression Inventory, Second Edition (BDI-II; Beck, Steer, & Brown, 1996) is a 21-item self-administered questionnaire assessing depression symptoms. For each item, there are four statements of increasing severity of depression symptoms. Scores range from 0–63 with higher scores indicative of greater severity of depression. The total score was used in analyses.

A subset of the participants (n = 49) were also administered the Mattis Dementia Rating Scale, Second Edition (DRS-2; Jurica, Leitten, & Mattis, 2001). The DRS-2 is another cognitive screening test that evaluates five cognitive domains (attention, initiation/perseveration, construction, conceptualization, and memory) and has been utilized to screen for cognitive impairment. Administration of the test is typically longer than the MoCA or MMSE, taking approximately 10–15 minutes in healthy older adults and 30 to 45 minutes in severely impaired individuals. The DRS-2 has excellent validity properties (Matteau et al., 2011; Matteau, Simard, Jean, & Turgeon, 2008).

Classification of participants

A diagnosis of MCI was assigned in this study by VA clinicians based on the criteria set forth by Winblad et al. (2004). All impairments noted 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). 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 disorders were available through medical records and/or formally diagnosed during the clinical evaluation according to DSM-IV-TR criteria.

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 neuropsychological 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 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. Preliminary analyses demonstrated little clinical meaning in analyzing such a heterogeneous group; therefore, results from analyses with the “Other” group were not reported. Thus, the focus will be on only three groups: MCI, No Diagnosis, and MDD.

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 [67.72, 73.72], p = .018) and those diagnosed with MCI (M = 69.60, SD = 7.63, 95% CI [67.85, 71.35], p = .111). Education level did not differ based on clinical diagnosis (F(3, 143) = 0.34, p = 0.80).

Analyses

The statistical analyses were performed using the SPSS software version 19.0. Associations between the SLUMS and neuropsychological measures were analyzed with Pearson correlations. The following variables were not significantly correlated with the SLUMS scores: age; education; potential medication effects; and histories of substance abuse, concussions, sleep disturbance, depression, cardiac disease, or post traumatic stress disorder (PTSD). 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-II both positively correlated with the SLUMS.

The diagnostic utility of the SLUMS was evaluated using ROC curve analyses. 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 AUC was calculated to estimate the overall diagnostic utility of the SLUMS. Positive predictive values (PPV) and negative predictive values (NPV) were also calculated.

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 adequate sample to detect a significant area under the curve (AUC; Wilcoxon-Mann-Whitney test) using a large adequate sample size was adequate for the planned analyses.

**Results**

**Data screening**

The data were examined for univariate and multivariate outliers, skewness, kurtosis, linearity, and multicollinearity. 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, with no concerns 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 if SLUMS performance differed significantly amongst individuals with no diagnosis, MCI, or MDD. The ANOVA revealed significant differences in performance amongst the groups (F(3, 143) = 25.54, p < 0.001). Post-hoc Bonferroni comparisons demonstrated that individuals diagnosed with MCI (M = 22.43, SD = 3.32) performed significantly poorer than those diagnosed with MDD (M = 26, SD = 3.50; p < 0.001, 95% CI [−5.80, −1.26], Cohen’s d = 1.05) and those with no diagnosis (M = 26.24, SD = 2.70, p < 0.001, 95% CI [−6.02, −1.61], Cohen’s d = 1.26). Interestingly, those with no diagnosis and those diagnosed with MDD did not perform significantly different (p = 1.00, 95% CI [−2.47, 3.04], Cohen’s d = 0.08).

Pearson product-moment correlations were conducted amongst SLUMS scores, demographic variables, and neuropsychological test variables (Table 2). The SLUMS total score was 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 range of significant correlation coefficients between the SLUMS and individual neuropsychological tests ranged from .25 (p < .01; Trails A) to .46 (p < .01; CVLT-2 Long Delay Free Recall). The SLUMS was also significantly correlated with the BDI-II (r(53) = 0.32, p = 0.02).

**ROC curve analyses 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 or b) MDD. Sensitivity and specificity values are provided for all cutoff scores in Table 3.

**MCI vs. No diagnosis**

The AUC discriminating between those diagnosed with MCI and no diagnosis 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 PPV was 88% and the NPV was 55%.

**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%, Specificity = 70%, PPV = 89%, NPV = 44%).

Overall, the SLUMS is adequate for discriminating between those with MCI and those with no diagnosis or MDD. Its low negative predictive values when compared to the positive predictive values suggest that a score above 24 or 25 does not necessarily indicate that an individual would not meet criteria for MCI following more comprehensive evaluation; however, a practitioner may feel more confident referring a patient for a neuropsychological assessment if they score below a 24 or 25 on the SLUMS.

In brief, there was some evidence to support the use of the SLUMS as a screening measure to discriminate between MCI and no diagnosis or MDD. Effect sizes (Table 4) were generally large (ranged from 1.05 to 1.26). The effect size for the difference between MCI
### Table 2. Correlations between SLUMS score, demographic factors, and neuropsychological test scores.

<table>
<thead>
<tr>
<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>
<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>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td>−.14</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>.11</td>
<td>.16</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>.44**</td>
<td>−.14</td>
<td>.02</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>.30**</td>
<td>.24**</td>
<td>.62**</td>
<td>.10</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></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>.29**</td>
<td>.23**</td>
<td>.63**</td>
<td>.11</td>
<td>1.00**</td>
<td>—</td>
<td></td>
<td></td>
<td></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>.31**</td>
<td>.22**</td>
<td>.63**</td>
<td>.10</td>
<td>.99**</td>
<td>.99**</td>
<td>—</td>
<td></td>
<td></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 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>—</td>
<td></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>.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>—</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>.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>—</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>.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>—</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>.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>
<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>.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>
<td>.37**</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td>.42**</td>
<td>.47**</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>.43**</td>
<td>.41**</td>
<td>.32**</td>
<td>.23*</td>
<td>.32**</td>
<td>.25**</td>
<td>.24**</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td>.23**</td>
<td>.37**</td>
<td>.24**</td>
<td>.24**</td>
<td>—</td>
<td></td>
<td></td>
<td></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>
<td>.26**</td>
<td>.38**</td>
<td>.43**</td>
<td>.24**</td>
<td>.58**</td>
<td>—</td>
<td></td>
<td></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>
<td>−.02</td>
<td>−.07</td>
<td>.04</td>
<td>.06</td>
<td>−.01</td>
<td>−.01</td>
<td>—</td>
<td></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>
<td>.11</td>
<td>−.10</td>
<td>−.03</td>
<td>.24**</td>
<td>.24**</td>
<td>.02</td>
<td>0</td>
<td>.64**</td>
</tr>
<tr>
<td>19. BDI-II</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>
<td>.27</td>
<td>.26</td>
<td>.23</td>
<td>.22</td>
<td>.33*</td>
<td>.24</td>
<td>.45**</td>
<td>.42**</td>
</tr>
</tbody>
</table>

Note. 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; CVLT-II = California Verbal Learning Test, Second Edition; LD = Long Delay; BVMT-R = Brief Visual Memory Test-Revised; ACBS = Anticholinergic Burden Scale; BDI-II = Beck Depression Inventory, Second Edition.

**p < 0.01. *p < 0.05.
and No Diagnosis was large and greater than for MCI and MDD. The effect sizes, AUCs, and diagnostic indicators for the SLUMS are considered good for a cognitive screening measure, with only NPV being less than desirable.

Discussion

The current study evaluated the psychometric properties of the SLUMS in a sample of veterans referred to a MCI clinic for evaluation. The SLUMS appears to be tapping into similar cognitive abilities as the comprehensive neuropsychological assessment, including simple attention and working memory, learning and memory, visuospatial skills, language and aspects of executive functioning; as the total score significantly correlated with all individual tests (with the exception of Trails B) and the DRS-2, suggesting that the SLUMS is a valid screening indicator of a patient’s performance on a more comprehensive neuropsychological battery.

The lack of correlation with Trails B is surprising, but may suggest that the SLUMS does not adequately assess cognitive flexibility and may be more effective in conjunction with Trails A and B; additional research assessing the incremental validity is recommended.

Two ROC curve analyses were conducted to evaluate the diagnostic accuracy of the SLUMS. Performance on the SLUMS was able to adequately discriminate between individuals diagnosed with MCI and those given no diagnosis or those diagnosed with MDD. The AUCs were fair to good, which suggests the SLUMS is a suitable measure to inform practitioners whether or not they should refer for comprehensive neuropsychological testing. Specifically, a score of less than 24 or 25 suggests the need for further neuropsychological assessment.

In addition to providing further validation for the SLUMS in a population with known or suspected MCI, the current study raised some interesting questions regarding further research. Most interesting was the differing cut off scores between those with MCI and No Diagnosis, and MCI and MDD, with an optimum cutoff score of 25 when discriminating between MCI and No Diagnosis, but a cutoff score of 24 when differentiating between MCI and MDD. Thus, it does appear that the SLUMS is negatively impacted by depression, thus assessing for current symptoms of affective distress will be important by health care providers in order to prevent false positive errors.

A limitation of the SLUMS is that it may not be free from the effect of culture, language, or quality of education, which is a weakness for its use in diverse populations. The test offers different cut scores for individuals with low education, but the SLUMS may

Table 3. Diagnostic indicators for the SLUMS.

<table>
<thead>
<tr>
<th>Cutoff score (≤)</th>
<th>Sensitivity</th>
<th>1-Specificity</th>
<th>Sensitivity</th>
<th>1-Specificity</th>
<th>Sensitivity</th>
<th>1-Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLUMS 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>
</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>
</tr>
</tbody>
</table>

Note. SLUMS = Saint Louis University Mental Status; PPV = Positive Predictive Value; NPV = Negative Predictive Value; MCI = Mild Cognitive Impairment; Dx = Diagnosis.
be sensitive to factors associated with race/ethnicity, given the significant difference in performance between European Americans and African Americans. As such, practitioners should use caution when using this measure with multicultural patient populations. Differences in performance were not accounted for by years of formal education, but given the correlation with the WTAR and the SLUMS score, it is possible that a measure of word reading may be thought of as an estimate of the quality of a person’s education (Manly, Jacobs, Touradji, Small, & Stern, 2002). Particular items that demonstrated disparities in the current sample (with African Americans performing lower than European Americans), included mental calculation and differences on the story memory task. Further research is needed to determine if there are more appropriate cut scores for African Americans that correctly distinguish between MCI and No Diagnosis.

A second limitation, specific to the current study, was the lack of inclusion of any women in the study sample. In our sample, participants were concurrently referred to the clinic were included in the data sample, and it was largely by chance that none were women. While the VA has traditionally been more male, increasing numbers of Veterans are women, and further study is needed to ensure that there are no gender differences in performance on the SLUMS. Finally, although there was adequate power to perform the analyses in the study as noted in the Methods section, this study utilized a relatively small sample size for ROC analyses, which often rely on much larger sample sizes for reliable estimates of sensitivity.

In sum, the SLUMS is a brief cognitive screener that appears to provide an adequate estimation of cognitive functioning that can help providers determine who should be referred for additional neuropsychological assessment. The measure adequately discriminates between individuals diagnosed with MCI and either MDD or no diagnosis; however, practitioners should be aware that there is a risk of false negatives (i.e., someone performing in the “Normal” range who has MCI).

Acknowledgement

Research was conducted at the Atlanta VA Medical Center.

ORCID

Suzanne Penna  http://orcid.org/0000-0002-1460-5735

References


adherence in outpatients with heart failure. *Heart & Lung, 41*(6), 572–582. doi:10.1016/j.hlnt.2012.06.001


