ABBREVIATED AND EXPANDED FORMS OF THE MONTREAL

COGNITIVE ASSESSMENT FOR DEMENTIA SCREENING

APPROVED BY THE SUPERVISORY COMMITTEE

_____

C. Munro Cullum, Ph.D., ABPP (Committee Chair)

_____

Linda S. Hynan, Ph.D.

_____

Laura H. Lacritz, Ph.D., ABPP

_____

Heidi C. Rossetti, Ph.D.

_____

Myron F. Weiner, M.D.

DEDICATION

In loving memory of Justis Lane Langford, who exemplified the true meaning of love, selflessness, and sacrifice for others. He touched the lives of all who knew him, and his passionate pursuit of the things he believed in will always inspire me.

ABBREVIATED AND EXPANDED FORMS OF THE MONTREAL COGNITIVE ASSESSMENT FOR DEMENTIA SCREENING

by

DANIEL KEVIN HORTON

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center

Dallas, Texas

August, 2015

ACKNOWLEDGEMENTS

I would like to express my gratitude for all the support and guidance I have received throughout my graduate endeavors. First and foremost, I thank God, to whom I attribute all of my accomplishments. I am continually reminded that all things are possible through Him. I want to thank my mentor, Dr. Munro Cullum, for believing in me and challenging me in all aspects of my professional development. Dr. Cullum helped me redefine my ideas about what is personally achievable, and I can only hope to inspire others in the same way throughout my career.

I want to thank each member of my dissertation committee for their unique contributions. I am so grateful for Dr. Linda Hynan's assistance with my development of a new skillset and her patience throughout the process. I also thank Dr. Laura Lacritz, who taught me so much about the field of neuropsychology through her instruction over the years. Dr. Myron Weiner was a source of encouragement who helped me improve my writing skills through astute critiques of my work. I also want to thank Dr. Heidi Rossetti for her support and contribution of knowledge on my dissertation subject area.

I am so appreciative for the guidance provided by Dr. Lana Harder throughout my final year of graduate school. In addition to her assistance with transferring my assessment skills to pediatric populations, Dr. Harder was especially supportive throughout the process of applying for postdoctoral

fellowships. I want to thank my incredible wife, Sarah, for her limitless patience, understanding, and encouragement over the past few years. She is such a blessing and I am ever inspired by her humility and work ethic. Finally, I want to thank my parents for their unwavering support. They have served as the primary role models in my life and I could have never achieved this without them.

July 2015

ABBREVIATED AND EXPANDED FORMS OF THE MONTREAL COGNITIVE ASSESSMENT FOR DEMENTIA SCREENING

Daniel Horton, Ph.D.

The University of Texas Southwestern Medical Center, 2015

Supervising Professor: C. Munro Cullum, Ph.D., ABPP

Cognitive screening is becoming increasingly important as the general population ages and the prevalence of dementia rises. However, popular cognitive screening tools have been criticized for their insensitivity to subtle cognitive impairment, poor specificity, excessive administration time, and/or questionable methods of test development. The Montreal Cognitive Assessment (MoCA) is a cognitive screening instrument growing in popularity which has demonstrated increased sensitivity to mild cognitive impairment (MCI), but takes roughly 10-15 minutes to administer and was developed without an empirically-driven item selection process. We devised two studies to address common limitations of cognitive screening tools using the MoCA.

The aim of Study 1 was to create a short form of the MoCA (SF-MoCA) including only the items found to be most sensitive to MCI and Alzheimer disease

(AD) and compare the diagnostic classification accuracy of the SF-MoCA to the Mini-Mental State Examination (MMSE) and standard MoCA. Results revealed delayed recall, orientation, and serial subtraction items to be most useful in differentiating the diagnostic groups. Overall, diagnostic accuracy of the SF-MoCA was superior to the MMSE and comparable to the standard MoCA, suggesting that some MoCA items do not add to the sensitivity of the instrument in these populations. Given the brevity and sensitivity of the SF-MoCA, we suggested this measure may be useful for early detection of cognitive impairment in primary care and other settings where evaluation time is limited.

Despite the advantages of the SF-MoCA, this tool only assesses three cognitive domains and may not be appropriate in settings where clinicians may want to efficiently assess additional domains affected in AD and MCI to gain a clearer picture of global functioning and assist in differential diagnosis. Therefore, we conducted a second study to determine if diagnostic accuracy of the SF-MoCA might be enhanced through the addition of several brief and well-validated neuropsychological measures shown to be sensitive to cognitive impairment. Results revealed that the addition of measures of processing speed, category fluency, and verbal recall resulted in an Expanded SF-MoCA with diagnostic classification accuracy superior to both the standard MoCA and SF-MoCA. Findings of these studies have implications for current cognitive screening procedures and techniques used to develop these tools.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF DEFINITIONS

AD – Alzheimer Disease

ADC – Alzheimer's Disease Center

AUC – Area Under the Curve

CERAD – Consortium to Establish a Registry for Alzheimer's Disease

CVLT-II – California Verbal Learning Test, Second Edition

MCI – Mild Cognitive Impairment

MIS – Memory Impairment Screen

MoCA – Montreal Cognitive Assessment

MMSE – Mini Mental State Examination

NC – Normal Control

NINDS-CSN VCI – National Institute of Neurologic Disorders and Stroke – Canadian Stroke Network Vascular Cognitive Impairment

NINCDS-ADRDA – National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association

NRI – Net Reclassification Improvement

ROC – Receiver Operating Characteristic

SF-MoCA – Short Form MoCA

SPSS – Statistical Package for the Social Sciences

WAIS – Wechsler Adult Intelligence Scale

SECTION II: STUDY 1

An Abbreviated Montreal Cognitive Assessment (MoCA) for Dementia Screening

Daniel K. Horton[1], Linda S. Hynan[1,2], Laura H. Lacritz[1,3], Heidi C. Rossetti[1], Myron F. Weiner[1], & C. Munro Cullum[1,3]

[1]Department of Psychiatry, [2]Department of Clinical Sciences (Biostatistics), [3]Department of Neurology and Neurotherapeutics

The University of Texas Southwestern Medical Center

Abstract

Objective: The Montreal Cognitive Assessment (MoCA) is a cognitive screening instrument growing in popularity, but few studies have conducted psychometric item analyses or attempted to develop abbreviated forms. We sought to derive and validate a short form MoCA (SF-MoCA) and compare its classification accuracy to the standard MoCA and MMSE in mild cognitive impairment (MCI), Alzheimer disease (AD), and normal aging.

Methods: 408 participants (MCI n=169, AD n=87, normal n=152) were randomly divided into derivation and validation sets. Item analysis in the derivation set identified most sensitive MoCA items. Receiver Operating Characteristic (ROC) analyses were used to develop cutoff scores and evaluate the diagnostic accuracy of the SF-MoCA, standard MoCA, and MMSE. Net Reclassification Improvement (NRI) analyses and comparison of ROC curves were used to compare classification accuracy of the three measures.

Results: Serial subtraction (Cramer's V=.408), delayed recall (Cramer's V=.702), and orientation items (Cramer's V=.832) were included in the SF-MoCA based on largest effect sizes in item analyses. Results revealed 72.6% classification accuracy of the SF-MoCA, compared with 71.9% for the standard MoCA and 67.4% for the MMSE. Results of NRI analyses and ROC curve comparisons revealed that classification accuracy of the SF-MoCA was comparable to the standard version and generally superior to the MMSE.

Conclusions: Findings suggest the SF-MoCA could be an effective brief tool in detecting cognitive impairment.

Introduction

As early detection of dementia becomes increasingly important, there is a growing need for quick and effective cognitive screening measures. The Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) was developed as an alternative to the popular Mini-Mental State Examination (MMSE; Folstein, Folstein & McHugh, 1975) and has shown sensitivity to mild cognitive impairment (MCI; Markwick, Zamboni, & de Jager, 2012; Smith, Gildeh, & Holmes, 2007). Although the MMSE and MoCA are relatively brief (i.e., 10-15 minutes administration time), their use in routine healthcare visits may still be limited due to time constraints (Iliffe et al., 2009; Tangalos et al., 1996). Therefore, many clinicians have a limited ability to quickly screen for cognitive impairment and may have difficulty determining when to refer for further evaluation when cognitive concerns are raised.

Abbreviated forms of the MMSE have been developed to address these issues by discarding less useful items and retaining those with higher discriminative value while maintaining classification accuracy. Three-word recall and orientation items on the MMSE are best able to discriminate demented from healthy older adults (Braekhus, Laake, & Engedal, 1992; Galasko et al., 1990), and many abbreviated forms of the MMSE and other cognitive screening tests have included these items to maintain sensitivity (e.g. Schultz-Larsen, Lomholt, & Kreiner, 2007). Haubois et al. (2011) created a 6-point MMSE short form

including only immediate and delayed recall items which showed similar sensitivity (89.5% versus 90.0%) and increased specificity (85.4% versus 75.5%) compared to the standard MMSE when screening for dementia. Despite inherent performance variability of three-word recall (Cullum, Thompson & Smernoff, 1993), the importance of delayed recall tasks is also demonstrated in studies which show that the Mini-Cog, comprised of three-word recall and a clock drawing task, is effective for detecting dementia (Borson, Scanlan, Chen, & Ganguli, 2003; Borson, Scanlan, Watanabe, Tu, & Lessig, 2005).

The MoCA is growing in popularity as a brief cognitive screening measure that is freely available (www.mocatest.org). In addition to the advantage of increased sensitivity to MCI, the MoCA is available in multiple languages and several alternate forms are available in English for the purposes of repeat evaluations. Despite these advantages, there have been limited psychometric analyses of MoCA items and few attempts to develop abbreviated versions. An abbreviated MoCA composed of orientation items, immediate and delayed recall items, and a phonemic fluency task was proposed by the neuropsychology working group of the National Institute of Neurologic Disorders and Stroke – Canadian Stroke Network Vascular Cognitive Impairment (NINDS-CSN VCI) for cognitive screening of patients with vascular disease (Hachinski et al., 2006). Freitas, Simões, Alves, Vicente, and Santana (2012) found it to be psychometrically sound and able to discriminate 34 patients with vascular

dementia from 34 controls (area under the curve = .936). However, this is the only publication to date that examines the classification accuracy of an abbreviated MoCA, and is limited by its non-empirical item selection and primary focus on a relatively small sample of patients with vascular dementia.

Given that Alzheimer disease (AD) is the most common form of dementia and is often preceded by MCI, an abbreviated MoCA effective in identifying and discriminating these populations would be useful. An abbreviated MoCA would also be valuable in addressing the criticisms of lengthy administration time in some settings and the inclusion of items that do not contribute to overall sensitivity. The purpose of the current study was to derive a short-form MoCA (SF-MoCA) and compare its classification accuracy to the standard MoCA and MMSE in distinguishing patients with MCI, AD, and cognitively normal controls (NC). We hypothesized that orientation and recall items would best distinguish between diagnostic groups given that similar items have demonstrated such utility and are commonly included in cognitive screening tools (e.g., Borson, Scanlan, Brush, Vitaliano, & Dokmak, 2000; Hachinski et al., 2006; Haubois et al., 2011).

Method

Participants

The sample was comprised of 408 participants (169 MCI, 87 AD, 152 controls) who underwent neurodiagnostic evaluation at the University of Texas Southwestern Medical Center Alzheimer's Disease Center between January 2012

and February 2014 in compliance with institutional regulations. Clinical diagnoses of *possible* and *probable* AD were made according to NINCDS/ADRDA criteria and MCI was diagnosed using Petersen criteria (Petersen, 2004) based upon multidisciplinary consensus review of neurologic, psychiatric, and neuropsychological data, in addition to patient- and informant-based reports. The MMSE was administered as part of the uniform data set of the National Alzheimer's Coordinating Center, and the MoCA was added to our center's evaluations in 2012 to explore its utility. MMSE scores within the AD group ranged from 9 to 29 with an overall mean of 21.2, reflecting a mild level of dementia.

The total sample (76.7% Caucasian) was divided into two groups: 1) the derivation set included a random selection of approximately 75% of cases (126 MCI, 67 AD, 124 NC) via the "select cases" function in SPSS 21, and 2) the remaining cases (43 MCI, 20 AD, 28 NC) comprised the validation set. Demographic characteristics and test scores for each set are displayed in Table 1.

Statistical Analyses

Derivation set

Item analyses of individual MoCA items were used to determine which items best differentiated the three diagnostic groups. Performance of each item was evaluated using Cramer's V with the expectation that the best items would have the largest effect sizes. The three items with the greatest effect sizes based

on item analyses were selected for the SF-MoCA. One-way ANOVA was used as a preliminary test to examine differences among the groups on the SF-MoCA. To allow comparison of effect sizes across measures, one-way ANOVAs of the standard MoCA and MMSE were also conducted.

Receiver Operating Characteristic (ROC) analyses were used to evaluate and compare the sensitivity and specificity of the SF-MoCA, standard MoCA, and the MMSE, first in the detection of cognitive impairment. Specifically, we used these analyses to examine the classification accuracy of each measure in distinguishing the NC group from a combined group of MCI + AD participants ("cognitively impaired" group). We also used ROC analyses to examine the accuracy of each measure in discriminating the AD group from the remainder of the sample for the purpose of cutoff score development. Combining diagnostic groups allowed us to use the entire sample with each ROC analysis and develop cutoff scores for both MCI and AD groups. Cutoff scores were developed for each measure based on sensitivity, specificity, accuracy, and the perpendicular distance between the selected point and the line of equality (Riffenburgh, 2006). ROC curves for the SF-MoCA were compared to the MMSE using a method developed by Hanley and McNeil (1983). Comparisons were made between the standard and short forms of the MoCA in the accurate detection of cognitive impairment using Net Reclassification Improvement (NRI) analyses (Pencina, D'Agostino, D'Agostino, and Vasan, 2007), which quantify differences in classification rates

between two measures or models. We employed this method to compare the standard and short forms of the MoCA instead of ROC comparisons due to overlap of items and increased statistical power of this analysis.

Validation set

Similar analyses were conducted in the validation set to support initial findings. ROC analyses were conducted for each measure to examine accuracy in distinguishing the NC group from the cognitively impaired group. The SF-MoCA and MMSE were compared in the accurate detection of cognitive impairment via statistical comparison of ROC curves, while this comparison was made between the standard and abbreviated MoCA versions using NRI analyses. Classification accuracy of cut scores selected in the derivation set was evaluated by correct predictions via crosstabs.

Additional analyses

To further explore the discrimination of groups using the SF-MoCA, we conducted additional analyses in both derivation and validation sets. ROC analyses were performed using the predictions from logistic regression in models that included the SF-MoCA and adjusted for age, sex, and education. The added predictive value of the risk-adjusted models was evaluated using NRI analyses. This method was used to determine if the inclusion of demographic measures into the final ROC models would significantly improve discriminative value.

ROC analyses were used in the validation set to test the classification accuracy of each measure in discriminating the AD group from the remainder of the sample. Comparisons of ROC curves were conducted to compare the SF-MoCA and MMSE in distinguishing these groups in both derivation and validation sets. Classification accuracy of the standard and SF-MoCA were compared in discriminating these groups using NRI analyses in both derivation and validation sets. Assumptions for all analyses were reviewed. In cases where assumptions were violated, we compared the findings of non-parametric tests to the parametric tests and in all cases the results were similar; we have reported the results of the parametric tests.

Results

Derivation Set

Item analysis of individual MoCA items in the derivation set revealed that serial subtraction (Cramer's V = .338) and delayed recall (Cramer's V = .489) were the best individual items at distinguishing between the NC and MCI groups. In discriminating between MCI and AD, item analyses showed that serial subtraction (Cramer's V = .408), delayed recall (Cramer's V = .702), and orientation items (Cramer's V = .832) best distinguished groups. These three items were selected for the SF-MoCA due to larger effect sizes (Cramer's V > .300) relative to other items and, therefore, greater accuracy in effectively distinguishing the three diagnostic groups. The inclusion of these items resulted in

a maximum SF-MoCA score of 14. Item analyses when differentiating between NC and AD groups were not used in determining most sensitive items due to limited variability among effect sizes for items.

The SF-MoCA performed well in distinguishing the clinical samples. One-way ANOVA revealed a statistically significant difference between the three groups in terms of MMSE scores (F (2, 313) = 246.2; $p < .001$), with a large effect size ($\eta_p^2 = .611$). Similar findings were observed for standard MoCA scores (F (2, 314) = 254.5; $p < .001$; $\eta_p^2 = .618$). Slightly larger differences between groups were seen on the SF-MoCA relative to the other screening measures (F (2, 314) = 333.04; $p < .001$; $\eta_p^2 = .680$).

In ROC analyses, the abbreviated and standard forms of the MoCA showed similar areas under the curve (AUC) when distinguishing the NC group from the cognitively impaired group as a whole (AUC = .86 vs .88). The SF-MoCA also demonstrated a similar AUC to the standard MoCA when differentiating patients with AD from the rest of the sample (AUC = .96 vs .93). The SF-MoCA showed a slightly larger AUC than the MMSE when differentiating controls from those with cognitive impairment (AUC = .86 vs .79). The SF-MoCA demonstrated a similar AUC to the MMSE when distinguishing patients with AD from the remainder of the derivation set (AUC = .96 vs .95). Figures 1 and 2 display ROC curves for each measure in the derivation set.

A significant difference was detected between ROC curves of the SF-MoCA and the MMSE when distinguishing between healthy controls and the cognitively impaired group ($z = 3.13$; $p < .01$), with the SF-MoCA outperforming the MMSE. NRI analyses showed no significant differences between the accuracy of the standard and short forms of the MoCA when distinguishing between these groups (NRI = 0.02; $p = .73$).

ROC analyses revealed an optimal cutoff score of <12 on the SF-MoCA to detect MCI and a cut point of <9 to detect AD. These cutoff scores resulted in an overall classification accuracy of 72.6% when classifying participants as NC, MCI, or AD. Cut scores of <26 and <20 on the standard MoCA were used to classify MCI and AD, respectively. The standard MoCA exhibited a classification accuracy of 71.9% when using these selected cutoff scores. Finally, classification accuracy of the MMSE was found to be 67.4% when using cut points of <29 and <25 to detect MCI and AD.

Validation Set

Results in the validation set were similar to findings in the derivation set. The effect sizes from one-way ANOVAs revealed largest group differences on the SF-MoCA compared to the other measures (F (2, 88) = 51.9; $p < .001$; $\eta_p^2 = .541$). The abbreviated and standard forms of the MoCA showed highly similar AUCs when differentiating healthy controls from the combined cognitively impaired group (AUC = .81 vs .83). Similar to results in the derivation set, the

AUC for the SF-MoCA was slightly larger than the MMSE when discriminating between these groups (AUC = .81 vs .78). Figure 3 displays ROC curves for each measure discriminating controls from cognitively impaired groups in the validation set.

When using cutoff scores selected in the derivation set, the SF-MoCA correctly classified 61.5% of participants as NC, MCI, or AD in the validation set. The standard MoCA exhibited similar classification accuracy, as derived cut scores resulted in 63.7% total accuracy. Finally, the MMSE correctly classified 57.8% of participants when using cutoff scores selected in the derivation set. Table 2 presents the classification accuracy of each measure using the derived cut points.

NRI analyses in the validation set revealed no significant differences in classification accuracy of the standard and SF- MoCA when differentiating controls from the cognitively impaired group (NRI = 0.04; p = .68). Similarly, no significant difference was found when comparing ROC curves of the SF-MoCA and MMSE when differentiating these groups (z = 0.76; p = .45).

Additional Analyses

In the derivation set, NRI analyses revealed that accuracy was not significantly improved when the ROC models for the SF-MoCA accounted for age, sex, and education when differentiating controls from the cognitively impaired group (NRI = .02; p = .29) or AD from the rest of the sample (NRI =

.03; p = .21). In light of these results, demographic variables were not incorporated into the final ROC models. Additional NRI analyses showed no significant differences between the accuracy of the standard and short forms of the MoCA when distinguishing AD from the rest of the derivation set (NRI = -0.03; p = .42). Comparison of ROC curves revealed no significant difference between the SF-MoCA and MMSE when differentiating between these groups in the derivation set (z = 0.70; p = .48).

In terms of differentiating patients with AD from the remainder of the validation set, the standard and SF- MoCA showed highly similar AUCs (AUC = .93 vs .91). Similar to results in the derivation set, the AUC for the SF-MoCA was slightly larger than the MMSE when discriminating between these groups (AUC = .93 vs .88). These ROC analyses are depicted in Figure 4. NRI analyses revealed no significant differences in the classification accuracy of the standard and short forms of the MoCA when discriminating patients with AD from the rest of the validation set (NRI = -0.10; p = .43). Finally, no significant difference was detected when comparing ROC curves of the SF-MoCA and MMSE when distinguishing between these groups in the validation set (z = 1.48; p = .14).

Discussion

Brief cognitive screening instruments are important in both clinical and research settings, particularly as there is a push to include cognitive checkups as part of healthcare wellness visits (Borson et al., 2007; Brayne, Fox, & Boustani,

2007). This was reinforced by the 2010 Patient Protection and Affordable Care Act, which requires providers to "detect any cognitive impairment" as part of the annual wellness visit for Medicare recipients. In addition to the limitation of relatively low sensitivity to subtle cognitive impairment, most current cognitive screening examinations are criticized for relatively lengthy administration times and include test items that do not enhance sensitivity. To address these issues, we sought to develop a short form of the MoCA which could decrease administration time while maintaining classification accuracy.

Orientation, word recall, and serial subtraction items were the best discriminators between diagnostic groups and were included in the SF-MoCA. We expected orientation and recall items to distinguish groups with higher accuracy than other items due to their sensitivity to cognitive impairment and inclusion in other brief cognitive screening instruments (e.g., Borson et al., 2000; Folstein et al., 1975; Hachinski et al., 2006; Haubois et al., 2011). While short forms of common cognitive screening tests have not typically included a serial subtraction component, this task is often used in mental status examinations (e.g., Strub & Black, 1985) and added to discriminability in the current sample. These results are consistent with an item analysis of the MMSE by Schultz-Larsen, Kreiner, and Lomholt (2007) which found that demented participants most commonly lost points on "orientation to time and place," "three-object recall," and "serial sevens."

The SF-MoCA consists of orientation, recall, and serial subtraction items, resulting in a maximum score of 14 points. The diagnostic groups showed statistically significant differences in scores on each of the three screening measures examined, with largest differences observed on the SF-MoCA in both the derivation and validation sets. The SF-MoCA showed larger AUCs than the MMSE in both samples when discriminating healthy controls from those with cognitive impairment, as well as patients with AD from the rest of the sample. Moreover, comparison of AUCs in the derivation set using the method of Hanley and McNeil (1983) revealed a significant difference between the SF-MoCA and MMSE when differentiating healthy controls from those with cognitive impairment. Although this finding was not replicated in the validation set, results of these ROC analyses support the ability of the SF-MoCA to distinguish between diagnostic groups with equal or better accuracy than the MMSE. These findings were further corroborated following the development of cutoff scores for each measure. Using cut points of <12 for MCI and <9 for AD, the SF-MoCA had an overall classification accuracy of 72.6% in the derivation set and 61.5% in the validation set. In contrast, the MMSE showed a classification accuracy of 67.4% in the derivation set and 57.8% in the validation set using cutoff scores of <29 for MCI and <25 for AD. It should be noted that our MMSE cutoff of <25 to detect AD is 1 point higher than the traditional cut score of <24 most commonly cited in the literature to detect dementia, although this slightly higher score is not

surprising given the above-average education of our groups. Overall, the classification accuracy of the SF-MoCA was superior to the MMSE in the current sample, as evidenced by better performance in each of the analyses conducted. Given that the MMSE has been widely used and even considered the "gold standard" of cognitive screening instruments (Boustani, Peterson, Hanson, Harris, & Lohr, 2003; Landi et al., 2000), these results support the use of the SF-MoCA for cognitive screening.

The SF-MoCA also performed well when compared to the standard MoCA. In both the derivation and validation sets, the SF-MoCA showed a slightly larger AUC than the standard MoCA when distinguishing patients with AD from the rest of the sample. However, AUCs of the standard MoCA in both samples were slightly larger than the SF-MoCA when discriminating controls from those with cognitive impairment. Using cutoff scores of <26 and <20 to classify MCI and AD, the standard MoCA demonstrated classification accuracies of 71.9% and 63.7% in the derivation and validation sets, respectively, compared to 72.6% and 61.5% for the SF-MoCA. It is also worth noting that the cut scores derived for the standard MoCA in the current study are consistent with the cutoff score of <26 originally proposed by Nasreddine et al. (2005) to detect cognitive impairment. NRI analyses showed that differences in the classification accuracy of these measures in each sample were small, non-significant, and unlikely to be clinically relevant.

These findings furthermore suggest that some MoCA items do not add appreciably to the clinical sensitivity of the standard version and that similar classification rates can be achieved with an abbreviated version. This is not surprising given that not all items on most omnibus cognitive screening instruments contribute to sensitivity. Rossetti, Lacritz, Cullum, and Weiner (2011) found that some MoCA items were useful in detecting cognitive impairment, while other items were rarely missed. Consequently, the SF-MoCA may be a comparably accurate, more efficient alternative to the standard form.

One strength of this study is the use of statistically-based selection criteria for items in the SF-MoCA. This contrasts with the techniques involved in developing many cognitive screening instruments wherein items are selected based on clinical judgment and/or loosely on prior research (e.g., Folstein et al., 1975; Hachinski et al., 2006; Nasreddine et al., 2005). Given that we employed a statistically-based method to develop the SF-MoCA, this specific combination of items should maximally distinguish similar diagnostic groups, although replication in larger and more heterogeneous samples will be needed to verify this. Another strength of this study involves the validation of results found in the derivation set. Whereas replication of our findings in future studies is needed to provide additional support for our conclusions, we were able to demonstrate that our findings generalized to a separate sample. Although classification accuracy of the SF-MoCA was decreased in the validation set, this is a common result of

testing a model that optimally fits the original data (Kohavi, 1995). Furthermore, the classification accuracy of the MMSE and standard MoCA were also reduced in the validation set, suggesting that this finding was not specific to the SF-MoCA.

This study has several limitations. The education level of our overall sample was relatively high, and the results may not generalize to individuals with less education. Additionally, as with the development of many abbreviated test versions, the SF-MoCA was derived from administration of the standard version. Hence, the short form was not administered as a unique test, and it is possible that the changes in administration timing might influence performance on some items. Along these same lines, a standard procedure for the administration of the SF-MoCA has not been addressed, and total administration time has not yet been determined.

Overall, the classification accuracy of the SF-MoCA was comparable to the standard version and generally superior to the MMSE. This suggests that the SF-MoCA could be an effective and efficient brief tool for raising suspicion of or detecting gross cognitive impairment, although the results of any brief cognitive screening test should not be interpreted in isolation or used to make a clinical diagnosis. Directions for future research include the development of a standardized protocol for administration of the SF-MoCA and evaluation of its utility and sensitivity in various clinical settings and populations.

Acknowledgements: This work was supported by the National Institutes of Health (grant number P30Ag12300-19). All authors declare no conflicts of interest in this work.

Table 1

Clinical Characteristics of the Derivation and Validation Sets

	NC	MCI	AD
Derivation Set			
n	124	126	67
Age	69.6 (7.9)	69.4 (7.7)	74.4 (8.1)
Education	15.3 (2.6)	14.5 (2.8)	15.2 (2.7)
% Female	67	52	37
% Caucasian	86.3	63.5	83.6
SF-MoCA	12.7 (1.2)	10.8 (1.9)	4.7 (3.3)
MoCA	27.0 (2.1)	23.0 (3.5)	14.1 (6.0)
MMSE	28.9 (1.1)	27.7 (1.8)	20.4 (4.9)
Validation Set			
n	28	43	20
Age	70.0 (7.0)	70.2 (6.1)	74.7 (7.9)
Education	16.2 (2.1)	15.2 (3.0)	15.3 (2.5)
% Female	46	56	35
% Caucasian	92.9	65.1	80.0
SF-MoCA	12.7 (1.6)	11.3 (1.6)	7.1 (2.9)
MoCA	26.7 (2.2)	24.2 (2.8)	17.0 (5.7)

MMSE	29.0 (0.8)	27.6 (2.3)	23.8 (3.5)

Note. NC = Normal Control; MCI = Mild Cognitive Impairment; AD = Alzheimer Disease; MoCA = Montreal Cognitive Assessment; SF-MoCA = Short-Form Montreal Cognitive Assessment; MMSE = Mini Mental State Examination. Data are presented as means (standard deviation) unless otherwise noted.

Table 2

Cutoff Scores and Diagnostic Accuracy of SF-MoCA, Standard MoCA, and MMSE in Derivation and Validation Sets

			Total Accuracy	
Measure	MCI cutoff	AD cutoff	Derivation	Validation
SF-MoCA	< 12/14	< 9/14	72.6%	61.5%
MoCA	< 26/30	< 20/30	71.9%	63.7%
MMSE	< 29/30	< 25/30	67.4%	57.8%

Note. MCI = Mild Cognitive Impairment; AD = Alzheimer Disease; MoCA = Montreal Cognitive Assessment; SF-MoCA = Short-Form Montreal Cognitive Assessment; MMSE = Mini Mental State Examination. Total accuracy is based on correct classification of participants as NC, MCI, or AD.

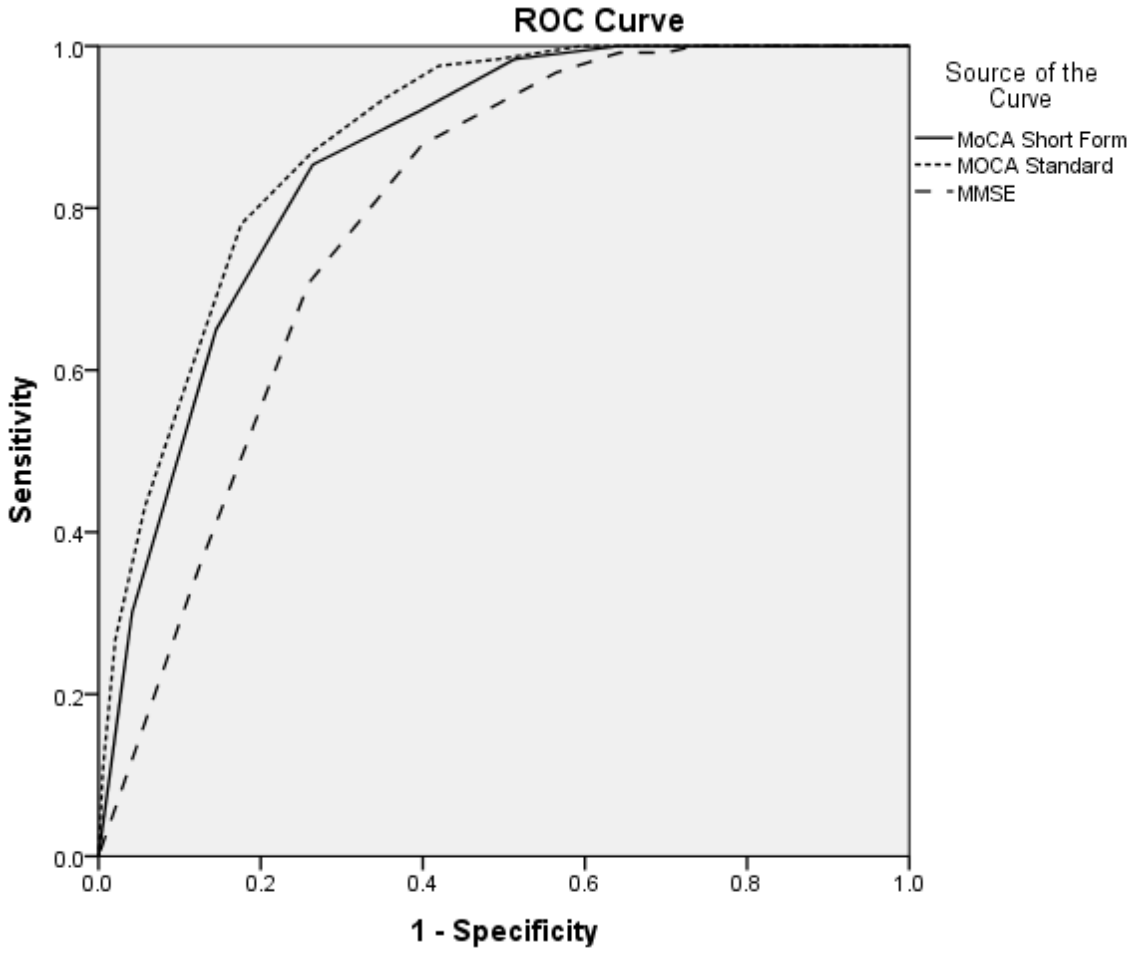

Figure 1. Receiver Operating Characteristic (ROC) curves for the SF-MoCA, Standard MoCA, and MMSE in differentiating controls from the cognitively impaired group (MCI+AD) in the derivation set. MCI = Mild Cognitive Impairment; AD = Alzheimer Disease; MoCA = Montreal Cognitive Assessment; SF-MoCA = Short-Form Montreal Cognitive Assessment; MMSE = Mini Mental State Examination.

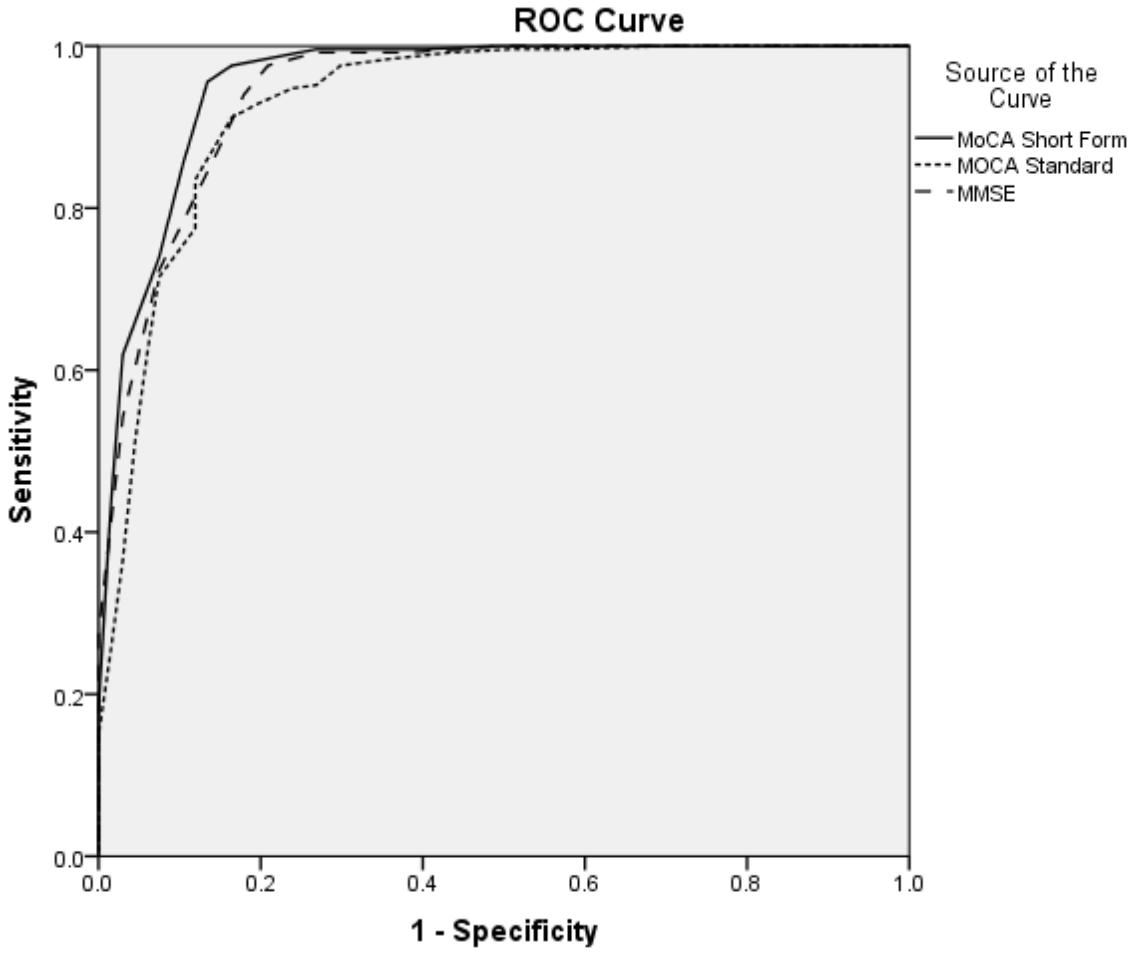

Figure 2. Receiver Operating Characteristic (ROC) curves for the SF-MoCA, Standard MoCA, and MMSE in differentiating AD from the remainder of the derivation set (NC+MCI). NC = Normal Control; MCI = Mild Cognitive Impairment; AD = Alzheimer Disease; MoCA = Montreal Cognitive Assessment; SF-MoCA = Short-Form Montreal Cognitive Assessment; MMSE = Mini Mental State Examination

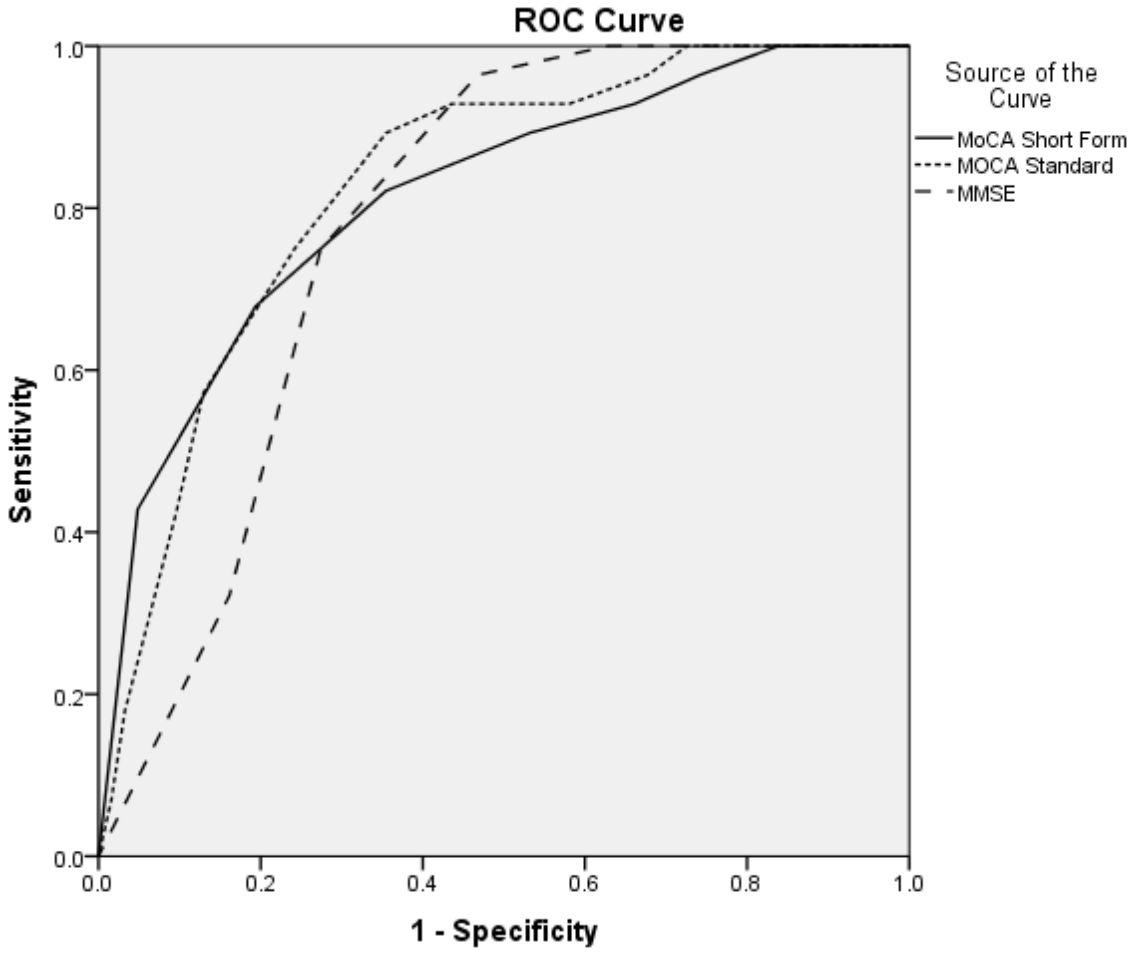

Figure 3. Receiver Operating Characteristic (ROC) curves for the SF-MoCA, Standard MoCA, and MMSE in differentiating controls from the cognitively impaired group (MCI+AD) in the validation set. MCI = Mild Cognitive Impairment; AD = Alzheimer Disease; MoCA = Montreal Cognitive Assessment; SF-MoCA = Short-Form Montreal Cognitive Assessment; MMSE = Mini Mental State Examination.

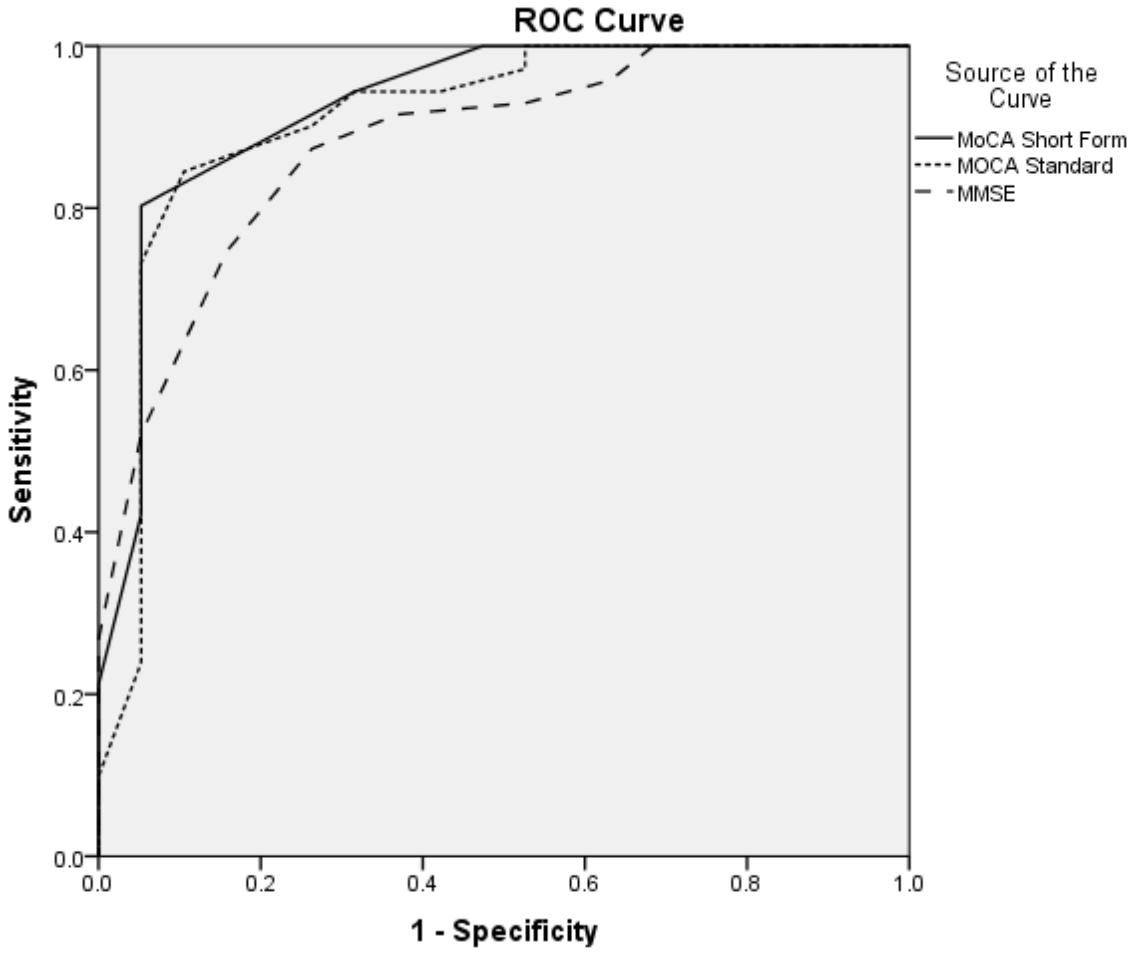

Figure 4. Receiver Operating Characteristic (ROC) curves for the SF-MoCA, Standard MoCA, and MMSE in differentiating AD from the remainder of the validation set (NC+MCI). NC = Normal Control; MCI = Mild Cognitive Impairment; AD = Alzheimer Disease; MoCA = Montreal Cognitive Assessment; SF-MoCA = Short-Form Montreal Cognitive Assessment; MMSE = Mini Mental State Examination.

SECTION III: STUDY 2

An Expanded Montreal Cognitive Assessment (MoCA) for the Detection of Mild Cognitive Impairment and Alzheimer Disease

Daniel K. Horton[1], Linda S. Hynan[1,2], Laura H. Lacritz[1,3], Heidi C. Rossetti[1], Myron F. Weiner[1], & C. Munro Cullum[1,3]

[1]Department of Psychiatry, [2]Department of Clinical Sciences (Biostatistics), [3]Department of Neurology and Neurotherapeutics

The University of Texas Southwestern Medical Center

Abstract

Objective: In previous work we developed a short form of the Montreal Cognitive Assessment (SF-MoCA) and found it to have classification accuracy comparable to the standard MoCA and superior to the MMSE in differentiating healthy controls and participants with mild cognitive impairment (MCI) and Alzheimer disease (AD). The aim of the present study was to enhance the sensitivity and item domain coverage of the SF-MoCA by supplementing it with brief and well-validated neuropsychological tasks known to be sensitive to impairment.

Methods: 408 participants (MCI n=169, AD n=87, normal n=152) were randomly divided into derivation and validation sets. Candidate supplementary measures from a variety of cognitive domains were identified based on brevity (i.e., 2-4 minutes) and sensitivity to MCI and/or AD. Effect sizes of independent-samples t-tests were calculated in the derivation set to determine the most useful supplementary measures for differentiating diagnostic groups. Two methods of scoring the screening battery (Expanded SF-MoCA) were explored to ensure maximum accuracy of the final model. Diagnostic classification accuracy of the Expanded SF-MoCA was compared to the standard MoCA and SF-MoCA via statistical comparison of Receiver Operating Characteristic (ROC) curves and Net Reclassification Improvement (NRI) analyses.

Results: Examination of effect sizes for candidate supplementary measures revealed Digit-Symbol Coding, Vegetable Fluency, and Trial 1 of the CERAD

word list to be most useful in distinguishing diagnostic groups. Results revealed 76.1% classification accuracy of the Expanded SF-MoCA, compared with 72.6% for the SF-MoCA and 71.9% for the standard MoCA. Results of NRI analyses and ROC curve comparisons revealed diagnostic accuracy of the Expanded SF-MoCA to be generally superior to the SF-MoCA and standard MoCA. Conclusions: Findings suggest the Expanded SF-MoCA may be a useful screening instrument for the detection and characterization of MCI and AD.

Introduction

Advances in healthcare and medicine have led to substantial increases in life expectancy over the past century (Kinsella & Wan, 2008). As the population ages and prevalence of dementia rises, cognitive screening instruments have become vital in the early detection of gross cognitive impairment in both clinical and research settings (Ismail, Rajji, & Shulman, 2010). Cognitive screening measures are commonly used by neuropsychologists to quickly assess gross cognitive functioning and track changes in global cognitive impairment over time (Salmon & Lange, 2001). There has also been demand to include cognitive screening as part of the medical workup in primary care and hospital settings (Borson et al., 2007; Brayne, Fox, & Boustani, 2007; National Institute for Health and Clinical Excellence and Social Care Institute for Excellence, 2006). In addition to use in clinical settings, these tools are often employed as an initial step of the diagnostic process in epidemiological studies and clinical treatment trials (Prince et al., 2013; Aarsland, Andersen, Larsen, & Lolk, 2003; Ott, Breteler, van Harskamp, Stijnen, & Hofman, 1998; Rogers, Farlow, Doody, Mohs, & Friedhoff, 1998; Rösler et al., 1999). Despite widespread use in a variety of settings, brief cognitive screening tools suffer from a number of limitations.

A major criticism of many cognitive screening instruments pertains to inadequate clinical sensitivity. Specifically, many of these brief tests have been found to be insensitive to subtle or mild cognitive impairment (MCI), which has

been conceptualized as a transition period between norming aging and early dementia (Lonie, Tierney, & Ebmeier, 2009; Petersen, 2004; Ravaglia et al., 2005). Additionally, some screening tools are unable to effectively distinguish MCI from dementia (Mitchell, 2009). Specificity has also been cited as a limitation, as screening procedures can result in false-positive diagnoses resulting in unnecessary distress and expensive testing for the individual labeled as demented or cognitively impaired (Ismail et al., 2010; Smith, Gildeh, & Holmes, 2007). Some widely used cognitive screening instruments have also been criticized for being too lengthy and impractical to use in settings where time is limited. Despite a public health initiative to include cognitive checkups as part of healthcare wellness visits, primary care physicians rarely use cognitive screening tools given their need to evaluate multiple body systems in a brief period of time (i.e., 15 minutes; Iliffe et al., 2009; Tangalos et al., 1996).

In addition to aforementioned limitations, many cognitive screening measures have been developed without the use of an empirically-driven item selection process. Some cognitive screening tools are derived based solely on clinical intuition or loosely based upon prior research, while item selection methods of many common instruments are not presented or unclear (e.g., Buschke et al., 1999; Folstein et al., 1975; Nasreddine et al., 2005). Authors often attempt to select a broad range of items that assess an array of cognitive abilities, but this approach can lead to excessive administration time without substantially

enhancing diagnostic accuracy. In light of these limitations, it has been difficult for clinicians to establish a consensus on the most useful cognitive screening instruments (Jacqmin-Gadda, Fabrigoule, Commenges, Letenneur, & Dartingues, 2000).

The Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) is the most extensively used cognitive screening tool to date and has been considered by some to be the "gold standard" of these measures (Boustani, Peterson, Hanson, Harris, & Lohr, 2003; Landi et al., 2000; Shulman et al., 2006). Despite its widespread use since its publication in 1975, the item selection process was never described in detail and it is subject to all of the limitations previously described (Mitchell, 2009; Tangalos et al., 1996; Tombaugh & McIntyre, 1992).

The Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) is a cognitive screening test growing in popularity which was developed as an alternative to the MMSE with a goal of increased sensitivity to MCI. It is available as a free download (www.mocatest.org) in multiple languages and has been found to be valid in a variety of ethnic populations (Fujiwara et al., 2010; Lee et al., 2008; Rahman, Gaafary, & Mohamed, 2009). Additionally, the MoCA has been demonstrated to be effective in detecting MCI and Alzheimer disease (AD), with sensitivity values ranging from .81 to .84 when distinguishing MCI and .88 to .94 when discriminating AD from cognitively normal controls (Freitas,

Simões, Alves, & Santana, 2013; Roalf et al., 2013; Smith et al., 2007). However, similar to previous investigations which have shown that some MMSE items do not significantly add to the test's sensitivity (Galasko et al., 1990; Tombaugh & McIntyre, 1992), item analyses in Study 1 revealed that certain items are useful in distinguishing these groups while other items are relatively insensitive regardless of an examinee's cognitive ability. Specifically, results of Study 1 demonstrated that MoCA orientation, delayed recall, and serial subtraction items were best able to detect MCI and AD, while items purported as sampling the domains of language, abstraction, visuospatial, and executive functioning were not useful in distinguishing these groups. For this reason, we developed a short form of the MoCA (SF-MoCA) using the items which demonstrated diagnostic classification accuracy comparable to the standard version and superior to the MMSE when discriminating patients with MCI, AD, and cognitively normal controls (NC). Results revealed overall classification accuracies of 72.6% for the SF-MoCA, 71.9% for the standard MoCA, and 67.4% for the MMSE. Despite good classification accuracy and brevity achieved with the SF-MoCA (i.e., about 5 minutes), the diversity of items is limited to three cognitive domains, and it is possible that sensitivity/specificity might be enhanced with the addition of other brief and sensitive cognitive tasks that are not currently included in the MoCA.

Researchers have taken various approaches to enhance the accuracy of cognitive screening tools. Some have sought to achieve this through modifying

administration procedures or adding individual items to existing screening measures, which have shown to be efficient ways of modestly increasing sensitivity of screening measures such as the MMSE and Addenbrooke's Cognitive Examination (Mathuranath, Nestor, Berrios, Rakowicz, & Hodges, 2000; McDowell, Kristjansson, Hill, & Hebert, 1997; Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006; Teng & Chui, 1987). However, modifications and additional items are often presented by authors without clear rationale. Researchers have also employed the technique of supplementing screening measures with additional brief cognitive tests assessing various cognitive domains. This approach is advantageous in that cognitive domains are more thoroughly assessed through measures that have been empirically supported through previous research, thereby increasing sensitivity to potential areas of deficiency. Although combining measures increases administration time, it can be effective in increasing the diagnostic accuracy of screening procedures (Beinhoff, Hilbert, Bittner, Gron, & Riepe, 2005; Commenges et al., 1992; Jacqmin-Gadda et al., 2000; Ravaglia et al., 2005; Xu, Meyer, Thornby, Chowdhury, & Quach, 2002). In order to maximally distinguish diagnostic groups through this approach, it is imperative that sensitive supplementary measures which have been empirically validated are chosen for inclusion in a screening battery.

When developing any cognitive screening instrument, the item selection process is crucial to accomplish the purpose of the measure. Shulman (2000)

proposed several criteria for effective cognitive screening measures including brevity and sensitivity. In addition to these important qualities, ideal screening tools should be broad in their coverage of domains (Ismail, 2010). Although the SF-MoCA derived in our previous study demonstrated acceptable accuracy, additional measures selected for brevity and sensitivity could be useful in enhancing the diagnostic accuracy of the SF-MoCA while expanding the breadth of cognitive domains sampled.

The present study sought to enhance the sensitivity of the SF-MoCA in distinguishing between MCI, AD, and cognitively normal controls by using brief, well-validated and sensitive tasks that expand the scope of the SF-MoCA but do not add undue administration time. We also sought to compare the diagnostic accuracy of this screening battery to that of the standard and short forms of the MoCA. It was hypothesized that diagnostic accuracy of the SF-MoCA would be increased when supplemented by additional brief neuropsychological measures and that this screening battery would discriminate diagnostic groups with higher accuracy than the standard MoCA.

Method

Participants

Data were obtained from the database of the Alzheimer's Disease Center (ADC) at the University of Texas Southwestern Medical Center. The total sample was comprised of 408 participants (169 MCI, 87 AD, 152 cognitively healthy).

Clinical diagnoses of *possible* and *probable* AD were made according to NINCDS-ADRDA criteria and MCI was diagnosed using Petersen criteria (Petersen, 2004) based upon multidisciplinary consensus review of neurologic, psychiatric, and neuropsychological data, in addition to patient- and informant-based reports. Neuropsychological data were collected as part of routine visits to the ADC, along with other demographic, medical and history data. Participants were randomly divided into derivation and validation sets via the "select cases" function in SPSS 21 in order to evaluate models of screening battery development and cutoff scores derived in the derivation set. The derivation set included approximately 75% of cases (126 MCI, 67 AD, 124 NC) and the remaining participants were included in the validation set (43 MCI, 20 AD, 28 NC).

Test Selection

The screening battery (Expanded SF-MoCA) was comprised of the SF-MoCA and additional brief neuropsychological measures. Brief tests assessing a variety of cognitive domains (i.e., language, attention, processing speed, executive functioning) were identified from the ADC database as meeting two main criteria of being 1) brief (2-4 minutes) and 2) sensitive to MCI and/or AD based on prior literature: Boston Naming Test (15-item version; Lansing et al., 1999); Digit-Symbol Coding subtest of the Wechsler Adult Intelligence Scale (Wechsler 1997a; 2008); Trail Making Test (TMT) – Parts A and B (Army Individual Test Battery, 1944); Verbal Fluency (Category and Letter; Benton,

Hamsher, Sivan, 1994; Spreen & Benton, 1977); Word List Trial 1 of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological battery (Morris, Mohs, Rogers, Fillenbaum, & Heyman, 1988). Effect sizes of each measure were examined in the derivation set to determine the most useful supplementary tests when discriminating between NC and MCI groups, as well as MCI and AD. The top three measures with highest overall discriminatory power were selected for inclusion in the Expanded SF-MoCA in order to maximize diagnostic accuracy and coverage of cognitive skills while keeping administration time reasonable for screening purposes. If multiple measures of similar domain content (e.g., letter fluency and category fluency) showed comparable discriminatory power, only one of the tasks was included in order to minimize test administration time while maximizing unique factors in the final model. For this reason, phonemic fluency tasks were examined separately rather than in combination to determine most sensitive measures.

Data Analysis

Screening battery development

Two techniques were explored and compared to determine the optimal method of assigning scores to supplementary measures to maximize predictions for diagnostic group membership. The first technique involved the development of cutoff scores for MCI and AD for each supplementary measure in the derivation set based on sensitivity, specificity, accuracy, and the perpendicular

distance between the selected point and the line of equality in Receiver Operating Characteristic (ROC) analyses (Riffenburgh, 2006). Following the development of cutoff scores, raw scores were converted to unit weights by assigning participants scores of 0, 1, or 2 on each supplementary test based on the cutoff ranges in which the raw scores fell. Specifically, participants were assigned a score of 0 on a particular measure if the raw score fell below the AD cutoff, 1 if the score fell in the MCI range, and 2 if the raw score was above the cut point for cognitively normal controls. Adjusted scores for each supplementary measure were subsequently combined with the SF-MoCA score, and this sum comprised the total score for the first model of the Expanded SF-MoCA.

In the second model, logistic regression analyses were used in the derivation set to assess the influence of each measure in the model and guide the process of determining score weights for each supplementary test. Logistic regression models included the SF-MoCA and all supplementary measures as predictors. The first regression model assessed the classification of participants as cognitively normal or impaired (MCI+AD group), while the second model distinguished AD from the remainder of the sample (NC+MCI). We chose to combine diagnostic groups in this way for two main reasons. First, this allowed us to increase power by including the entire sample in each analysis. Additionally, the classification of normal versus cognitively impaired is consistent with typical cognitive screening needs in clinical settings. After logistic regression models

were conducted, the b weights for each predictor in both models were averaged to form "average" b weights for each measure which took into account the results of both regression models. Each of the average b weights for the supplementary measures were then divided by the average b weight for the SF-MoCA, and the raw scores for each supplementary test were multiplied by these quotients to determine adjusted raw scores for each supplementary measure. The adjusted raw scores for the supplementary measures were added to the SF-MoCA score to determine the total score for this model of Expanded SF-MoCA.

The two techniques of scoring the Expanded SF-MoCA were compared via ROC analyses in the derivation and validation sets to determine the optimal scoring model. ROC analyses distinguishing controls from the cognitively impaired group (MCI+AD), as well as AD from the rest of the sample (NC+MCI) were conducted for each of the total scores resulting from both battery development methods. The technique that demonstrated better performance by ROC analyses was to be used in the remaining analyses, although the unit-weighted model would be chosen if ROC results were similar given its relative simplicity.

Expanded SF-MoCA vs SF-MoCA

Cutoff scores for MCI and AD for the Expanded SF-MoCA total score were derived in the derivation set using the same method (Riffenburgh, 2006) described in the previous section in order to determine classification accuracy. In

contrast, classification accuracy of the SF-MoCA was determined using cutoff scores derived in Study 1 using the same derivation set. Net Reclassification Improvement (NRI) analyses, which quantify improvement in classification accuracy when additional information or tasks are added to a particular measure, were conducted in both samples to compare the Expanded SF-MoCA to the SF-MoCA in differentiating controls, MCI, and AD. NRI analyses are appropriate when comparing these two measures because the SF-MoCA is embedded in the Expanded SF-MoCA and *improvement* in classification accuracy is being measured. Comparisons of ROC curves were also conducted in derivation and validation sets using the method described by Hanley and McNeil (1983) to determine if there were significant differences between the measures when distinguishing between normal controls and the cognitively impaired group (MCI+AD), as well as AD and the remainder of the sample (NC+MCI).

Expanded SF-MoCA vs Standard MoCA

Classification accuracy of the standard MoCA was determined in both samples using cut scores from Study 1. Statistical comparisons of ROC curves in both the derivation and validation sets were used to determine if there were statistically significant differences between the measures in terms of diagnostic accuracy.

Results

Screening Battery Development

Examination of effect sizes for candidate supplementary measures revealed Digit-Symbol Coding, Category Fluency (vegetables), and Trial 1 of the CERAD word list to be most useful overall in distinguishing the diagnostic groups in the derivation set. A complete listing of effect sizes for all potential supplementary measures considered is provided in Table 1.

The two methods of scoring the screening battery were highly similar in terms of diagnostic accuracy. When distinguishing between controls and the cognitively impaired group (i.e., MCI + AD), the unit-weighted model (model 1) had an area under the curve (AUC) of .89, compared to .88 for the logistic regression model (model 2). Additionally, the unit-weighted and logistic regression models demonstrated identical AUCs when discriminating between AD and the remainder of the derivation set (AUC = .96 vs .96). These results were confirmed in the validation set, as both models demonstrated AUCs of .85 when distinguishing controls from the cognitively impaired group in the validation set. Similarly, the unit-weighted model demonstrated a comparable AUC to the logistic regression model when distinguishing the AD group from the remainder of the validation set (.93 vs .94).

To further demonstrate parity between the models of scoring, intraclass correlation coefficients were calculated in both derivation and validation sets. The models were very strongly correlated in all conditions, with correlation coefficients of .98 and .99 in both samples. In light of these results, we elected to

proceed with the unit-weighted model given its relative simplicity.

Expanded SF-MoCA vs SF-MoCA

Cutoff scores of <17/20 and <9/20 were determined to be optimal for the classification of MCI and AD, respectively, in the derivation set using the Expanded SF-MoCA. The Expanded SF-MoCA demonstrated overall classification accuracies of 76.1% in the derivation set and 74.7% in the validation set based on these cutoff scores. By contrast, cut points of <12/14 and <9/14 were used to classify MCI and AD using the SF-MoCA based on findings from Study 1 in the derivation set, resulting in overall classification accuracies of 72.6% in the derivation set and 61.5% in the validation set.

Results of NRI analyses in the derivation set revealed no statistically significant difference between classification accuracy of the Expanded SF-MoCA and the SF-MoCA when distinguishing controls from the cognitively impaired group (NRI = 0.04; $p = .24$) or AD from the remainder of the derivation set (NRI = -0.09; $p = .07$). However, a statistically significant difference was detected between classification accuracy of the Expanded SF-MoCA and the SF-MoCA when distinguishing between controls and the cognitively impaired group in the validation set (NRI = 0.21; $p < .01$), with the Expanded SF-MoCA outperforming the SF-MoCA. There was no significant difference between the measures in terms of classification accuracy when differentiating AD from the remainder of the validation set (NRI = -0.16; $p = .22$).

Comparison of ROC curves revealed that the Expanded SF-MoCA performed significantly better than the SF-MoCA when distinguishing between controls and the cognitively impaired group in the derivation set ($z = 3.83$; $p < .01$). However, there was no significant difference between ROC curves when differentiating AD from the remainder of the derivation set ($z = 0.60$; $p = .55$). In the validation set, there were no significant differences between ROC curves of the measures when distinguishing controls from the cognitively impaired group ($z = 1.91$; $p = .06$) or AD from the remainder of the validation set ($z = 0.35$; $p = .72$).

Expanded SF-MoCA vs Standard MoCA

Cutoff scores of <26/30 and <20/30 were used to classify participants as MCI or AD using the standard MoCA based on findings from Study 1 in the derivation set, resulting in overall classification accuracies of 71.9% in the derivation set and 63.7% in the validation set (compared with 76.1% and 74.7% using the Expanded SF-MoCA). Cutoff scores and overall classification accuracies of each measure in both samples are displayed in Table 2. Comparison of ROC curves in the derivation set revealed no significant difference between the Expanded SF-MoCA and the standard MoCA when distinguishing controls from the cognitively impaired group ($z = 1.15$; $p = .25$). However, the Expanded SF-MoCA significantly outperformed the standard MoCA when distinguishing the AD group from the remainder of the derivation set ($z = 2.29$; $p < .05$). In the validation set, there were no significant differences between the measures when

differentiating controls from the cognitively impaired group (z = 1.07; p = .29) or AD from the remainder of the validation set (z = 0.87; p = .39). A complete listing of ROC results for each measure in both samples is displayed in Table 3.

Discussion

Brief cognitive screening measures are increasingly important in both clinical and research settings, but suffer a number of limitations. The MoCA is becoming more frequently used for early detection of cognitive impairment given that it is freely available and has demonstrated increased sensitivity to MCI relative to other popular screening measures such as the MMSE. Despite the widespread use of the MoCA and its relative advantages over other brief screening tools, it was revealed in Study 1 that not all MoCA items contribute to the test's sensitivity and comparable diagnostic accuracy can be achieved with an abbreviated version of the test (SF-MoCA). Although the SF-MoCA should prove particularly useful in primary care settings given its brevity and ability to detect gross cognitive impairment, it assesses a limited number of cognitive domains and may not be appropriate for settings where more detailed cognitive information is desired. The present study was aimed at enhancing the clinical utility of the SF-MoCA through supplementing this abbreviated form with several brief neuropsychological tasks in order to maximize diagnostic accuracy and provide additional clinical information regarding cognitive status without adding undue administration time.

Comparison of effect sizes revealed Digit-Symbol Coding, Category Fluency, and Trial 1 of the CERAD Word List to be the most useful brief neuropsychological tasks in distinguishing diagnostic groups. These findings make sense given that impairments in verbal episodic memory, semantic knowledge, verbal fluency, and processing speed are common in AD and MCI, although cognitive profiles vary by subtype. (Cullum & Lacritz, 2009; Petersen, 2004; Twamley & Bondi, 2004). In light of these results, we chose to include Digit-Symbol coding, Vegetable Fluency, and Trial 1 of the CERAD word list as supplementary measures in the Expanded SF-MoCA.

Two methods of scoring the Expanded SF-MoCA were compared to ensure that the final model would maximally distinguish diagnostic groups. Results revealed that a unit-weighted model performed comparably to a more complex model that adjusted scores of supplementary measures based on their predictive influence in logistic regression models. These findings are consistent with Wainer's (1976) finding that coefficients in linear regression models can be replaced with equal weights without loss in predictive accuracy. A recent study by Donohue et al. (2014) provides further support for this approach, as the authors found that a similar unit-weighted technique could be used to reliably detect cognitive decline in preclinical AD. Our decision to employ the unit-weighted model was also based on its simplicity, as this scoring approach requires little effort on the clinician's part to obtain a total score for the screening battery.

Details of scoring supplementary measures of the expanded MoCA are displayed in Table 4.

The Expanded SF-MoCA demonstrated higher overall classification accuracy than the SF-MoCA in both derivation and validation sets. Moreover, the Expanded SF-MoCA demonstrated equal or larger AUCs than the SF-MoCA in both samples when differentiating controls from those with cognitive impairment, as well as AD from the remainder of the sample. Results of NRI analyses and comparison of ROC curves provided further evidence of the superiority of the Expanded SF-MoCA relative to the SF-MoCA. These results demonstrate enhanced diagnostic accuracy of the SF-MoCA with the addition of three brief neuropsychological tasks without adding excessive administration time. These findings are not surprising given that we supplemented the SF-MoCA with well-validated tasks shown to be sensitive to MCI and AD. These findings are also consistent with studies demonstrating that diagnostic accuracy of screening measures can be significantly enhanced with the addition of brief and well-validated neuropsychological measures (Beinhoff et al., 2005; Commenges et al., 1992; Jacqmin-Gadda et al., 2000; Ravaglia et al., 2005; Xu et al., 2002).

The Expanded SF-MoCA also performed well compared to the standard MoCA in distinguishing diagnostic groups, as classification accuracy for the Expanded SF-MoCA was higher in both samples. Although reduction in classification accuracy is common when validating a model derived in a separate

sample, the reduction in accuracy of the Expanded SF-MoCA when tested in the validation set was negligible (i.e., 1.5%) compared to reductions of more than 10% for the standard MoCA and SF-MoCA. This stability of classification accuracy across samples using the Expanded SF-MoCA provides further support for the robustness of this combination of measures and method of scoring. Furthermore, the Expanded SF-MoCA outperformed the standard MoCA in all ROC analyses and a significant difference was detected between the Expanded SF-MoCA and the standard MoCA when differentiating AD from the remainder of the derivation set. Taken together, these findings suggest the Expanded SF-MoCA is superior to the standard MoCA in differentiating these diagnostic groups.

Results suggest we have developed a brief and sensitive screening battery with diagnostic accuracy superior to the SF-MoCA and standard MoCA when differentiating controls, MCI, and AD groups. Furthermore, the Expanded SF-MoCA is consistent with many of Shulman et al.'s (2000) criteria for an ideal cognitive screening battery, as this measure is brief, simple to administer and score, psychometrically robust, and taps a range of cognitive domains. The supplementary tasks included in the Expanded SF-MoCA may also assist clinicians in obtaining additional information regarding the neurocognitive status of patients beyond the use of an omnibus cut score. As mentioned previously, a

prominent strength of this screening battery is its brevity, as administration time should be less than or equal to the standard MoCA (i.e., 10 minutes or less).

Although a test form and administration protocol has not been developed for the expanded MoCA, we offer a suggested order of task administration summarized in Figure 1. We recommend the examiner begin with the two learning trials of the MoCA word list, followed by orientation items, serial subtraction, digit-symbol coding, and delayed recall of the MoCA word list. This administration results in a time delay between encoding and recall trials of the MoCA word list that is consistent with standard administration of the MoCA (i.e., approximately 5 minutes). Upon completion of the delayed recall trial, examiners may administer vegetable fluency followed by Trial 1 of the CERAD word list. We recommend administering the first trial of the CERAD word list at the conclusion of the screening battery so as to minimize potential interference between word lists. Although proactive interference may occur in some cases, recall errors such as intrusions do not affect the total score of the Expanded SF-MoCA. Although the total score does not account for intrusions, the occurrence of such errors from one list to the other may provide additional qualitative information that may be useful to the clinician.

A notable strength of this study is the statistically-driven selection of items for inclusion in the Expanded SF-MoCA. Given that tasks were selected for both the SF-MoCA as well as supplementary measures based on discriminatory power,

sensitivity of the Expanded SF-MoCA is maximized without adding unnecessary administration time. Additionally, we compared methods of scoring the Expanded SF-MoCA to ensure that our final model yielded the greatest sensitivity using the items we selected. Another strength is our use of a validation set to confirm results. Although replication of findings in future studies in various populations and settings would provide further validation for our findings, results showed that the diagnostic accuracy of the Expanded SF-MoCA was highly reliable across samples, providing further support for the robustness of the model.

There are several limitations of this study. First, the education level of the overall sample was relatively high (mean = 15.1 years), so use of cutoffs derived in this sample may lead to false positives in individuals with lower education levels. Refer to the Section IV (Integrated Conclusions) for additional discussion regarding the association between demographic variables and performance on the Expanded SF-MoCA, SF-MoCA, and standard MoCA. Another limitation is the derivation of the Expanded SF-MoCA based on analyses of retrospective data obtained from a larger battery of tests. Therefore, the Expanded SF-MoCA has not been administered as its own entity and unforeseen factors may influence test performance when tasks are administered in the proposed order. Generalizability of findings is also limited by our use of participants presenting for evaluation at an ADC, as this sample may not be representative of general clinical cases that present to other types of clinical settings. Finally, as with many cognitive

screening measures, the inclusion of multiple language-based measures can introduce cultural bias that may lead to false-positive diagnoses in non-native English speakers.

Diagnostic accuracy of the Expanded SF-MoCA was superior to the SF-MoCA and standard MoCA in almost all analyses conducted. Thus, this measure may be a useful instrument to assist in differential diagnosis of MCI and AD through quickly and effectively assessing cognitive domains commonly affected in these conditions. Future studies should explore the utility of the Expanded SF-MoCA in various clinical settings and populations, including potential cross-cultural effects. Along these lines, the sensitivity of the instrument should be examined in clinical settings where rates of dementia are lower than the context of an ADC. Future studies may also explore the utility of the Expanded SF-MoCA in staging cognitive impairment via longitudinal analyses.

Table 1

Effect Sizes (Cohen's d) for Candidate Supplementary Measures for the Expanded SF-MoCA

Measure	NC vs MCI	MCI vs AD	Mean
Digit-Symbol Coding	1.07	1.11	1.09
Vegetable Fluency	0.65	1.51	1.08
Animal Fluency	0.63	1.31	0.97
CERAD Word List Trial 1	0.88	1.02	0.95
Trail Making Test Part A	0.67	1.11	0.89
Trail Making Test Part B	1.05	0.58	0.81
Boston Naming Test (15 item)	1.00	0.55	0.77
"A" Fluency	0.79	0.70	0.75
"F" Fluency	0.63	0.71	0.67
"S" Fluency	0.74	0.45	0.60

Note. All effect sizes were calculated using Cohen's d. The "Mean" column represents the mean of effect sizes for each condition to give an overall estimate of discriminatory power.

Table 2

Cutoff Scores and Classification Accuracy of the Expanded SF-MoCA, SF-MoCA, and Standard MoCA in Derivation and Validation Sets

			Total Accuracy	
Measure	MCI cutoff	AD cutoff	Derivation	Validation
Expanded SF-MoCA	< 17/20	< 9/20	76.1%	74.7%
SF-MoCA	< 12/14	< 9/14	72.6%	61.5%
MoCA	< 26/30	< 20/30	71.9%	63.7%

Note. Total accuracy is based on correct classification of participants as NC, MCI, or AD.

Table 3

AUCs of the Expanded SF-MoCA, SF-MoCA, and Standard MoCA in the Derivation and Validation Sets

	Normal vs Cognitively Impaired	AD vs Remainder of the Sample
Derivation Set		
Expanded SF-MoCA	.89	.96
SF-MoCA	.86	.96
Standard MoCA	.88	.93
Validation Set		
Expanded SF-MoCA	.85	.93
SF-MoCA	.82	.93
Standard MoCA	.83	.92

Table 4

Raw Score Cutoffs for Supplementary Measures

Measure	Normal (2 points)	Mildly Impaired (1 point)	Severely Impaired (0 points)
Vegetable Fluency	> 12	7-12	< 7
Digit-Symbol Coding	> 46	31-46	< 31
CERAD Word List – Trial 1	> 5	4-5	< 4

Measure	**# of Possible Points**
1) MoCA Word List Learning (2 trials)	Unscored
2) MoCA Orientation items	____/6
3) MoCA Serial Subtraction	____/3
4) Digit-Symbol Coding	____/2
5) MoCA Word List Delayed Recall	____/5
6) Vegetable Fluency	____/2
7) CERAD Word List Trial 1	____/2
Total	**____/20**

Figure 1. List of tasks and associated scores in suggested order of administration for the Expanded SF-MoCA. MoCA = Montreal Cognitive Assessment; CERAD = Consortium to Establish a Registry for Alzheimer's Disease.

SECTION IV: INTEGRATED CONCLUSIONS

Overview

As the population ages, early detection of dementia and age-related cognitive impairment is increasingly important. Along these lines, there is a need for brief cognitive screening tools to provide a practical means to address this issue (Ismail et al., 2010). The MMSE has been the most widely employed cognitive screening measure for decades (Shulman et al., 2006), although limitations of this instrument are well-documented (Anthony, LeResche, Niaz, Von Korff, & Folstein, 1982; Faustman, Moses, & Csernansky, 1990; Tombaugh & McIntyre, 1992; Wind et al., 1997). Most importantly, the MMSE has been shown to have low sensitivity to mild or subtle cognitive impairment which may be indicative of future decline (Smith et al., 2007). The MoCA was developed to address this drawback and has become widely accepted by clinicians due to empirical support for its ability to detect MCI (Freitas et al., 2013; Nasreddine et al., 2005). Moreover, the MoCA is freely available in multiple languages and has demonstrated validity in a variety of ethnic populations (Fujiwara et al., 2010; Lee et al., 2008; Rahman, Gaafary, & Mohamed, 2009).

Despite these advantages of the MoCA, it was developed based on "clinical intuition" rather than an empirical process and administration time can be relatively lengthy (i.e., 10-15 minutes). Given that it was developed based on clinical intuition, we speculated that the MoCA may include some insensitive

items resulting in unnecessary administration time and that a short form may perform comparably to the standard version in detecting gross cognitive impairment. In Study 1, we conducted an item analysis of the MoCA to determine the most sensitive items to MCI and AD and created a short form of the MoCA (SF-MoCA) including these items. The diagnostic classification accuracy of the SF-MoCA was comparable to the standard version and superior to the MMSE in differentiating MCI, AD, and cognitively normal controls. These results demonstrated the importance of careful item selection through the use of statistics and/or consideration of task sensitivity based on previous research when developing cognitive screening instruments. Given its brevity and sensitivity to impairment, the SF-MoCA was proposed to be useful in primary care settings where time constraints are often a barrier to formal assessment of cognitive functioning.

While the SF-MoCA may be ideal in settings where time is restricted, it assesses a limited number of cognitive domains and may not be appropriate for assistance with differential diagnosis or when clinicians want a more thorough assessment of areas commonly affected in MCI or AD without spending significant time administering and scoring tasks. In Study 2, we proposed that sensitivity of the SF-MoCA may be enhanced with additional brief and well-validated neuropsychological tasks without adding undue administration time. To create this screening battery, we examined the utility of candidate measures

sampling a variety of cognitive domains and supplemented the SF-MoCA with the three measures with highest discriminatory power. Results of Study 2 revealed that the Expanded SF-MoCA differentiated diagnostic groups with superior accuracy to the SF-MoCA and standard MoCA. Given the superior performance of the Expanded SF-MoCA and the finding that the MoCA items excluded from the SF-MoCA did not enhance diagnostic accuracy, results suggest that a valid screening test such as the SF-MoCA can be improved only when supplemented with sensitive tasks. These findings further support the importance of careful item selection when developing a screening measure and attempting to maximize sensitivity while minimizing administration time.

Comparison with Other Cognitive Screening/Assessment Measures

Findings of Study 2 revealed a classification accuracy of 76.1% for the Expanded SF-MoCA when classifying participants as NC, MCI, or AD in the derivation set. Furthermore, the Expanded SF-MoCA revealed classification accuracy of 98% when distinguishing controls from AD and 78% when discriminating NC from MCI using proposed cutoff scores of < 9/20 and < 17/20. The SF-MoCA also performed well when distinguishing these groups, demonstrating an overall classification accuracy of 72.6% in the derivation set despite brief administration time (i.e., approximately 5 minutes). Using proposed cutoff scores of < 9/14 and < 12/14, the SF-MoCA correctly classified 98% of participants when distinguishing controls from AD and 70% of participants when

discriminating between controls and MCI. Accuracies of both measures when classifying participants into two diagnostic groups are displayed in Table 1 in Appendix C. A review of the literature examining the diagnostic accuracy of common cognitive screening instruments suggests the Expanded SF-MoCA may be among the most sensitive of these measures, although validation in additional samples would be needed to verify this.

Ismail et al. (2010) present a list of the most common cognitive screening tools used in various clinical settings. The Mini-Cog (Borson et al., 2000), a clock drawing task supplemented with a three-word recall task, is among one of the most widely used brief cognitive screening measures for the detection of dementia and has shown promise for screening in primary care settings (Lorentz, Scanlan, & Borson, 2002; Brodaty, Low, Gibson, & Burns, 2006). A validation study conducted by Borson et al. (2003) revealed sensitivity and specificity values similar to the MMSE for the detection of dementia (76% sensitivity and 89% specificity). The Mini-Cog has also demonstrated several advantages over the MMSE such as ease of administration and minimal language and education bias (Borson et al., 2005). Although administration time of the Mini-Cog is slightly shorter than the SF-MoCA, the SF-MoCA revealed superior sensitivity and specificity values when detecting dementia (96% and 99%). The Expanded SF-MoCA also revealed diagnostic accuracy superior to the Mini-Cog when detecting dementia (94% sensitivity and 100% specificity), although comparison between

the Mini-Cog and SF-MoCA may be more appropriate given similarity in administration time.

Another common cognitive screening measure discussed in the review by Ismail et al. (2010) is the Memory Impairment Screen (MIS; Buschke et al., 1999), a four-item delayed free and cued recall memory task. The MIS has several advantages, as it takes less than five minutes to administer and performance is not significantly influenced by age, sex, or education. When using the proposed cut score of 4 for classification of patients as demented or normal, the MIS demonstrated excellent specificity (96%). However, a sensitivity value of only 80% was reported for detection of dementia in the original validation study. Furthermore, this measure has not been validated as a screening instrument for MCI.

Although few cognitive screening instruments other than the MoCA have been validated for detection of MCI, DemTect (Kalbe et al., 2004) is a relatively brief measure (i.e., administration time of 8-10 minutes) that has shown promise in addressing this issue. This measure consists of five tasks: a word list encoding task, a number transcoding task, a verbal fluency task, digit span backward, and delayed recall of the word list. DemTect demonstrated correct classification of 81% of MCI patients and 85% of AD patients using proposed cutoffs of < 13/18 and < 9/18, respectively, in the original validation study conducted by the authors. However, a validation study by Larner (2007) showed an AUC of only .87 when

distinguishing controls from those with dementia (compared to .96 for both the SF-MoCA and Expanded SF-MoCA). AUCs of the SF-MoCA and Expanded SF-MoCA when classifying participants into two diagnostic groups are displayed in Table 2 in Appendix C.

For exploratory purposes, we also compared the diagnostic accuracy of the SF-MoCA and Expanded SF-MoCA to selected variables from the California Verbal Learning Test – Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000), a detailed measure of verbal learning and memory with high sensitivity to memory disorders (Ribeiro, Guerreiro, & DeMendonça, 2007; Zakzanis, Leach, & Kaplan, 1999). Results of these analyses are displayed in Table 3 in Appendix C. In short, the Expanded SF-MoCA exhibited greater accuracy than CVLT-II total learning, short-delay free recall, and long-delay free recall when differentiating controls from those with cognitive impairment, and the SF-MoCA demonstrated accuracy similar to these CVLT measures. All measures performed similarly when differentiating AD from the remainder of the sample. These findings provide further support for the utility of the SF-MoCA and Expanded SF-MoCA, as the CVLT has demonstrated high utility in detecting memory impairment in these conditions (Twamley & Bondi, 2004; Zakzanis et al., 1999).

Overall, these findings suggest that the SF-MoCA and Expanded SF-MoCA are comparable or superior to most popular cognitive screening tools in

differentiating MCI and AD from cognitively normal controls. Although additional research on these measures is warranted, our use of a validation set serves as preliminary confirmation of these findings.

Demographic Effects

The association between education and performance on cognitive testing is well-known (Bornstein & Suga, 1988). Education effects have been noted as a significant limitation of several popular cognitive screening measures, most notably the MMSE (Bravo & Hébert, 1997; Ismail et al., 2010; Uhlmann & Larson, 1991). The authors of the MoCA attempted to address this issue by adding one point to the total score for individuals with less than 12 years of education. However, subsequent research suggests the MoCA is subject to education effects regardless of this modification (e.g., see Rossetti et al., 2011). In order to explore the relationship between education and total score on the SF-MoCA, Expanded SF-MoCA, and standard MoCA, Pearson correlation coefficients were calculated within the control group and total sample. Details of these analyses are displayed in Appendix C. Findings revealed that the total scores on the SF-MoCA and Expanded SF-MoCA were less correlated with education (r = .15 and r = .22, respectively) than the standard MoCA (r = .27) in the control group. The SF-MoCA was least correlated with education in the total sample (r = .09), while the standard MoCA and Expanded SF-MoCA were equally correlated with education in the total sample (r = .18). When taken

together, these findings suggest the standard MoCA may be slightly more vulnerable to education influences than the SF-MoCA or Expanded SF-MoCA. Variation in dementia severity within the AD group may be a possible explanation for differences between correlations in the control group compared to the total sample.

Many areas of cognition tend to decline with age in healthy individuals, while some cognitive skills tend to remain stable or increase with age (Ardila & Rosselli, 1989). Correlations between age and performance on the standard MoCA, SF-MoCA, and Expanded SF-MoCA were calculated in the total sample to explore this relationship. Details of these analyses are displayed in Appendix C. Total scores on the standard MoCA and Expanded SF-MoCA showed highly similar correlations with age (r = -.26 and r = -.25, respectively), while the SF-MoCA total score showed a slightly higher, moderate correlation with age (r = -.30). These findings suggest the SF-MoCA may be slightly more vulnerable to the cognitive effects of aging. However, the differences between these correlations were small and unlikely to be clinically meaningful.

Sex is another demographic factor known to influence performance on cognitive tests. Analyses displayed in Appendix C revealed significant sex effects of the SF-MoCA and Expanded SF-MoCA in the control group. Specifically, women significantly outperformed men on both the SF-MoCA and Expanded SF-MoCA, while no significant difference was detected between sexes on the

standard MoCA. It should be noted that mean differences between sexes on the SF-MoCA were very small (i.e., less than 1 point). Mean scores on the standard MoCA, SF-MoCA, and Expanded SF-MoCA by sex are displayed in Table 4 in Appendix C. We posit that women outperformed men on the SF-MoCA and Expanded SF-MoCA mainly due to the highly verbal nature of most tasks selected, as women have been shown to outperform men on verbal measures (Lewin, Wolgers, & Herlitz, 2001; Weiss et al., 2003). Moreover, women may have had an advantage on the Digit-Symbol Coding and category fluency tasks, as Roivainen (2011) found that women tend to outperform men on processing speed tasks involving digits and letters, as well as rapid naming tasks. Despite sex differences on these measures, results of sex-specific ROC analyses using the standard MoCA, SF-MoCA, and Expanded SF-MoCA revealed that diagnostic accuracy of the Expanded SF-MoCA continued to be superior or equal to the standard MoCA and SF-MoCA when assessed in males and females separately. Also consistent with findings in Study 2, accuracy of the standard MoCA and SF-MoCA were similar in these analyses. Results of sex-specific ROC analyses are displayed in Table 5 in Appendix C. These results suggest that findings of Study 1 and Study 2 regarding comparison of accuracy between measures remain robust despite the influence of sex.

Although sex effects may be considered limitations of the SF-MoCA and Expanded SF-MoCA, sex differences on neuropsychological tasks are common

(Heaton, Miller, Taylor, & Grant, 2004; Levy & Heller, 1992). Analyses displayed in Appendix C also revealed significant differences between sexes on the CVLT-II total learning score, which is widely considered to have high utility in diagnosing AD and MCI. Despite sex differences on the Expanded SF-MoCA, this measure performed equal or better than standard MoCA when ROC analyses were conducted separately in males and females, suggesting this screening battery maintains utility regardless of patient sex.

Strengths and Limitations

A notable strength of these studies is our use of a relatively large validation set to confirm initial findings. Given that our results were generally consistent across samples, we can be confident in the utility of these measures in distinguishing diagnostic groups. Additionally, participants were carefully diagnosed by a multidisciplinary team using current diagnostic guidelines and procedures. Another strength is the statistically-based selection of items for inclusion in both the SF-MoCA and Expanded SF-MoCA. Items included in these measures and the respective domains assessed are consistent with common measures used in dementia assessment and literature on cognitive profiles of AD and MCI, providing further validation for this combination of measures.

One limitation of the current studies is the relatively high education level of the overall sample. Thus, we cannot be sure that our results will generalize to populations with lower levels of education. However, correlations between

education and performance on the standard MoCA, SF-MoCA, and Expanded SF-MoCA showed that performance on the SF-MoCA and Expanded SF-MoCA are less associated with education than the standard MoCA, suggesting that the SF-MoCA and Expanded SF-MoCA may be less prone to education effects. The highly verbal nature of the tasks included in the SF-MoCA and Expanded SF-MoCA could also be considered a limitation, as sex effects are commonly observed on verbal measures. Given that our sample was comprised of primarily Caucasian individuals, we cannot be sure that the cutoff scores derived in these studies will be directly applicable to non-Caucasian groups. However, exploratory analyses revealed that the classification accuracy of the SF-MoCA was similar in Caucasians and African-Americans (70.9% and 65.1%, respectively). Classification accuracy using the Expanded SF-MoCA was also similar between races (Caucasian = 74.5% and African-Americans = 79.1%). Finally, the SF-MoCA and Expanded SF-MoCA were developed based on analyses of retrospective data and were not administered separately from the standard MoCA and supplementary cognitive measures.

Future Directions

Future studies should attempt to validate the SF-MoCA and Expanded SF-MoCA in populations of diverse racial background and lower education levels, especially given the predominantly Caucasian makeup of the current sample. Future studies should also explore the accuracy of these measures when

administered as individual entities, as order of task administration may affect performance on certain items. Given the significant sex effects of the SF-MoCA and Expanded SF-MoCA, future studies may also explore sex-specific cutoff scores and their influence on diagnostic accuracy.

APPENDIX A

Aims and Hypotheses

Overall Aim: Investigate the utility of abbreviated and expanded forms of the MoCA in distinguishing cognitively normal controls, patients with MCI, and patients with AD.

Study 1

Aim 1: Determine the most useful individual MoCA items in distinguishing cognitively normal controls from MCI, as well as MCI from AD.

Hypothesis 1: Orientation and delayed recall will demonstrate highest utility in distinguishing the diagnostic groups.

Aim 2: Explore the utility of a short form of the MoCA (SF-MoCA) comprised of the most sensitive individual items determined in Aim 1.

Hypothesis 2a: NRI analyses will reveal no significant differences between the classification accuracies of the SF-MoCA and standard MoCA when distinguishing controls from the cognitively impaired group or the AD group from the remainder of the sample.

Hypothesis 2b: Statistical comparison of ROC curves will reveal no significant differences between the accuracy of the SF-MoCA and MMSE when distinguishing controls from the cognitively impaired group or the AD group from the remainder of the sample.

Study 2

Aim 1: Select brief and sensitive neuropsychological measures evaluating a variety of cognitive domains to supplement the SF-MoCA in order to create a novel cognitive screening battery (Expanded SF-MoCA).

Hypothesis 1: Category Fluency, Trails B, and Trial 1 of the CERAD word list will demonstrate largest effect sizes when differentiating diagnostic groups.

Aim 2: Determine an optimal method of adjusting raw scores of supplementary measures to incorporate them into a total score for the Expanded SF-MoCA to maximize group differences. Two methods of scoring will be compared.

Hypothesis 2: The Expanded SF-MoCA will discriminate between diagnostic groups with highest accuracy when the raw scores for supplementary measures are adjusted based on their predictive ability in logistic regression models.

Aim 3: Compare the diagnostic classification accuracy of the Expanded SF-MoCA to the SF-MoCA.

Hypothesis 3a: NRI analyses will reveal a significant improvement in diagnostic accuracy when the SF-MoCA is supplemented with additional brief measures. Specifically, NRI analyses will reveal a significant difference between the diagnostic accuracy of these

measures when distinguishing controls from the cognitively impaired group, with the Expanded SF-MoCA outperforming the SF-MoCA.

Hypothesis 3b: Statistical comparison of ROC curves will reveal a significant difference between the Expanded SF-MoCA and the SF-MoCA when distinguishing controls from the cognitively impaired group, with the Expanded SF-MoCA outperforming the SF-MoCA.

Aim 4: Compare the diagnostic classification accuracy of the Expanded SF-MoCA to the standard MoCA.

Hypothesis 4: Comparison of ROC curves will reveal a statistically significant difference between these measures when distinguishing controls from the cognitively impaired group, with the Expanded MoCA outperforming the standard MoCA.

Appendix B

Additional Background

Investigation of cognitive profiles in AD suggest that impairments in verbal episodic and semantic memory, word-finding, verbal fluency, visuospatial skills, and aspects of executive functioning are most common, with memory symptoms typically presenting earliest in the disease course (Cullum & Lacritz, 2009; Twamley & Bondi, 2004). Patients with MCI typically show deficiencies in similar domains, although cognitive profiles vary based on subtype (Petersen, 2004). While deficits in episodic memory are most commonly observed in patients with MCI (Dudas, Clague, Thompson, Graham, & Hodges, 2005; Wang & Zhou, 2002), declines in other areas such as attention, executive function, and semantic knowledge are often observed (Alladi, Arnold, Mitchell, Nestor, & Hodges, 2006; Arnáiz & Almkvist, 2003; Salmon, 2012).

The CERAD neuropsychological battery was developed to detect and quantify cognitive deficits associated with AD and therefore assesses all of the aforementioned domains. Given that verbal episodic memory impairment is the hallmark of AD and often impaired in MCI, the CERAD neuropsychological battery includes a word list with learning, recall, and recognition components which has been found to be sensitive to MCI and AD (Karrasch, Sinervä, Grönholm, Rinne, & Laine, 2005; Weintraub, Wicklund, & Salmon, 2012; Welsh, Butters, Hughes, Mohs, & Heyman, 1991). The CERAD word list is a 10-item

word list which is presented over three trials at a rate of 1 word every 2 seconds. The patient is asked to read each word as it is presented and recall the words immediately upon conclusion of the word list presentation. The examinee is then asked to recall these words after a short delay (i.e., 5 minutes) and subsequently recognize them from a list of target and distractor words. Although delayed recall has been found to be the most sensitive of these word list tasks (Welsh et al., 1991), some research has demonstrated differences between controls and AD, as well as controls and MCI on the first learning trial of this word list (Karrasch et al., 2005).

As mentioned previously, semantic knowledge and language skills are also frequently impaired in AD and MCI. Brief measures of verbal fluency (Spreen & Benton, 1977) are sensitive to deficits in these areas and are commonly used by neuropsychologists in the assessment of dementia (Cullum & Lacritz, 2009). In these tasks, examinees are asked to rapidly name as many words as possible that begin with a particular letter or belong to a given category within one minute. Although both types of verbal fluency measures are often used in dementia assessment, semantic (category) fluency is generally more impaired than phonemic (letter) fluency in both AD (Epker, Lacritz, & Cullum, 1999; Monsch et al., 1992, 1994) and MCI (Alladi et al., 2006).

The Digit-Symbol Coding subtest of the WAIS has also demonstrated sensitivity to cognitive deficits in a wide array of cognitive disorders including

AD (Zakzanis et al., 1999) and has proven useful in predicting cognitive decline in MCI and healthy elderly individuals (Arnáiz & Almkvist, 2003; Tabert et al., 2006). This task requires participants to rapidly draw symbols in numbered boxes according to the way the symbols are paired with numbers in a key at the top of the page. It is frequently used in dementia assessment given its brevity and wide range of cognitive abilities assessed including processing speed, visual scanning, working memory, and visual-motor coordination (Cullum & Lacritz, 2009).

Appendix C

Additional Analyses

Classification of Participants into Two Diagnostic Groups Using the Standard MoCA, SF-MoCA, and Expanded SF-MoCA

Accuracy was calculated when classifying participants as NC or MCI, MCI or AD, and NC or AD using the standard MoCA, SF-MoCA, and Expanded SF-MoCA in total sample. Results are displayed in Table 1.

Table 1

Accuracy when Classifying Participants into Two Diagnostic Groups Using the Standard MoCA, SF-MoCA, and Expanded SF-MoCA

	Control vs MCI	Control vs AD	MCI vs AD
Standard MoCA	73.6%	97.9%	81.2%
SF-MoCA	69.9%	98.0%	85.8%
Expanded SF-MoCA	77.9%	98.2%	89.2%

Differentiation of Individual Diagnostic Groups Using the Standard MoCA, SF-MoCA, and Expanded SF-MoCA

Diagnostic accuracies of the standard MoCA, SF-MoCA, and Expanded SF-MoCA were explored by distinguishing controls vs MCI, MCI vs AD, as well as controls vs AD in the total sample to determine if these measures maintain utility when diagnostic groups are not combined. Results of these ROC analyses are displayed in Table 2.

Table 2

AUCs of the Standard MoCA, SF-MoCA, and Expanded SF-MoCA When Differentiating Individual Diagnostic Groups

	Control vs MCI	Control vs AD	MCI vs AD
Standard MoCA	.82	.96	.87
SF-MoCA	.78	.97	.92
Expanded SF-MoCA	.84	.98	.93

Comparison of Diagnostic Accuracy of the CVLT, Standard MoCA, SF-MoCA, and Expanded SF-MoCA

We compared the accuracy of the standard MoCA, SF-MoCA, and Expanded SF-MoCA to the accuracy of the most sensitive CVLT variables via ROC analyses to further explore the utility of various versions of the MoCA. Results of ROC analyses are displayed in Table 3.

Table 3

AUCs of the Standard MoCA, SF-MoCA, Expanded SF-MoCA, and Most Sensitive CVLT Variables in the Total Sample

	Normal vs Cognitively Impaired	AD vs Remainder of the Sample
Standard MoCA	.85	.90
SF-MoCA	.83	.93
Expanded SF-MoCA	.88	.94
CVLT Total Learning	.84	.93
CVLT Short Delay Free Recall	.85	.94
CVLT Long Delay Free Recall	.83	.94

Correlations between Education and the Standard MoCA, SF-MoCA, and Expanded SF-MoCA

Pearson correlation coefficients were calculated to explore the relationship between years of education and total scores on the standard MoCA, SF-MoCA, and Expanded SF-MoCA in the control group and total sample. Results revealed the SF-MoCA total score to be least associated with education in both the control group ($r = .15$, $p = .06$) and total sample ($r = .09$, $p = .06$). The standard MoCA showed the highest correlation with education in the control group ($r = .27$, $p < .01$), followed by the Expanded SF-MoCA ($r = .22$, $p < .01$). In the total sample, the standard MoCA and Expanded SF-MoCA were equally correlated with education ($r = .18$, $p < .01$).

Correlations between Age and the Standard MoCA, SF-MoCA, and Expanded SF-MoCA

Pearson correlation coefficients were calculated to examine the association between age and performance on the standard MoCA, SF-MoCA, and Expanded SF-MoCA in the total sample. Performance on the Expanded SF-MoCA showed the lowest correlation with age (r = -.25), followed by the standard MoCA (r = -.26) and the SF-MoCA (r = -.30). All correlations were statistically significant (p < .01).

Sex Differences on the Standard MoCA, SF-MoCA, Expanded SF-MoCA, and CVLT in the Control Group

Independent samples t-tests were performed to investigate sex differences on the standard MoCA, SF-MoCA, Expanded SF-MoCA, and CVLT in the control group. Mean scores for each measure by group are provided in Table 4. Results revealed no significant difference between sexes on the standard MoCA (t (150) = 1.35, p = .18, two-tailed). A statistically significant difference was detected between groups on the SF-MoCA (t (94.53) = 2.47, p < .05, two-tailed), with women outperforming men. However, the effect size was relatively small (eta squared = .039) and the groups differed by less than 1 point. Results also revealed a statistically significant difference between the sexes on the Expanded SF-MoCA (t (88.74) = 3.79, p < .01, two-tailed), with women outperforming men. The magnitude of the difference between males and females on the Expanded SF-MoCA was moderate (eta squared = .088). No significant differences were detected between sexes on CVLT short-delay (t (81.41) = -0.60, p = .55, two-tailed) or long-delay free recall (t (81.65) = -0.71, p = .48, two-tailed). However, females scored significantly higher than males on CVLT total learning (t (143) = 4.45, p < .01, two-tailed), and the magnitude of these differences was large (eta squared = .12).

Table 4

Scores on the Standard MoCA, SF-MoCA, and Expanded SF-MoCA by Sex

	Males	Females
Standard MoCA	26.6 (2.1)	27.1 (2.1)
SF-MoCA	12.4 (1.5)	12.9 (1.2)
Expanded SF-MoCA	17.0 (2.2)	18.2 (1.6)

Note: Data are presented as means (standard deviation).

Sex-Specific ROC Analyses Using the Standard MoCA, SF-MoCA, and Expanded SF-MoCA

Given sex differences on the SF-MoCA and Expanded SF-MoCA, ROC analyses were conducted within male and female groups separately to examine sex-specific changes in diagnostic accuracy of these measures. All analyses were performed in the total sample and differentiated controls from those with cognitive impairment, as well as the AD group from the remainder of the total sample. Results are displayed in Table 5.

Table 5

Sex-Specific AUCs Using the Standard MoCA, SF-MoCA, and Expanded SF-MoCA

	Normal vs. Cognitively Impaired	AD vs. Remainder of the Sample
Males		
Standard MoCA	.85	.90
SF-MoCA	.81	.92
Expanded SF-MoCA	.85	.94
Females		
Standard MoCA	.87	.94
SF-MoCA	.85	.96
Expanded SF-MoCA	.90	.96

Incremental Validity of Expanded SF-MoCA Supplementary Measures

The utility of individual supplementary measures included in the Expanded SF-MoCA was further examined by adding the unit-weighted scores of the supplementary measures to the SF-MoCA individually rather than aggregating all scores into a total. The diagnostic accuracies of these three models were evaluated via ROC analyses in the total sample. ROC analyses distinguished controls from those with cognitive impairment, as well as AD from the remainder of the sample. ROC results using the standard MoCA, SF-MoCA, and Expanded SF-MoCA were also included to demonstrate changes in diagnostic accuracy with various combinations of items and measures. Results of these analyses are displayed in Table 6.

Table 6

Incremental Validity of Expanded SF-MoCA Supplementary Measures

	Normal vs Cognitively Impaired	AD vs Remainder of the Sample
Standard MoCA	.86	.92
SF-MoCA	.84	.94
SF-MoCA + Digit-Symbol Coding	.87	.94
SF-MoCA + Category Fluency (Vegetables)	.85	.95
SF-MoCA + CERAD Word List Trial 1	.86	.95
Expanded SF-MoCA	.88	.95

REFERENCES

Aarsland, D., Andersen, K., Larsen, J. P., & Lolk, A. (2003). Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Archives of Neurology*, *60*(3), 387-392.

Alladi, S., Arnold, R., Mitchell, J., Nestor, P. J., & Hodges, J. R. (2006). Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychological Medicine*, *36*(4), 507-516.

Ardila, A., & Rosselli, M. (1989). Neuropsychological characteristics of normal aging. *Developmental Neuropsychology*, *5*(4), 307-320.

Anthony, J. C., LeResche, L., Niaz, U., Von Korff, M. R., & Folstein, M. F. (1982). Limits of the 'Mini-Mental State'as a screening test for dementia and delirium among hospital patients. *Psychological Medicine*, *12*(02), 397-408.

Army Individual Test Battery: Manual of Directions and Scoring: Washington, DC, War Department, Adjutant General's Office, 1944.

Arnáiz, E., & Almkvist, O. (2003). Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. *Acta Neurologica Scandinavica 107*(Suppl. 179), 34-41.

Beinhoff, U., Hilbert, V., Bittner, D., Gron, G., & Riepe, M. W. (2004). Screening

for cognitive impairment: a triage for outpatient care. *Dementia and Geriatric Cognitive Disorders, 20*(5), 278-285.

Benton, A. L., Hamsher, K. D., & Sivan, A. B. (1994). *Multilingual aphasia examination: manual of instructions*. AJA Associations.

Bornstein, R. A., & Suga, L. J. (1988). Educational level and neuropsychological performance in healthy elderly subjects. *Developmental Neuropsychology, 4*(1), 17-22.

Borson, S., Scanlan, J., Brush, M., Vitaliano, P., & Dokmak, A. (2000). The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. *International Journal of Geriatric Psychiatry, 15*(11): 1021-1027.

Borson, S., Scanlan, J. M., Chen, P., & Ganguli, M. (2003). The Mini-Cog as a screen for dementia: validation in a population-based sample. *Journal of the American Geriatrics Society, 51*(10), 1451-1454.

Borson, S., Scanlan, J., Hummel, J., Gibbs, K., Lessig, M., & Zuhr, E. (2007). Implementing routine cognitive screening of older adults in primary care: process and impact on physician behavior. *Journal of General Internal Medicine, 22*(6), 811-817.

Borson, S., Scanlan, J. M., Watanabe, J., Tu, S. P., & Lessig, M. (2005). Simplifying detection of cognitive impairment: comparison of the Mini-Cog and Mini-Mental State Examination in a multiethnic

sample. *Journal of the American Geriatric Society, 53*(5), 871-874.

Boustani, M., Peterson, B., Hanson, L., Harris, R., & Lohr, K. N. (2003). Screening for dementia in primary care: a summary of the evidence for the US Preventive Services Task Force. *Annals of Internal Medicine, 138*(11), 927-937.

Braekhus, A., Laake, K., & Engedal, K. (1992). The Mini-Mental State Examination: identifying the most efficient variables for detecting cognitive impairment in the elderly. *Journal of the American Geriatrics Society, 40*(11), 1139-1145.

Bravo, G., & Hébert, R. (1997). Age-and education-specific reference values for the Mini-Mental and Modified Mini-Mental State Examinations derived from a non-demented elderly population. *International Journal of Geriatric Psychiatry,12*(10), 1008-1018.

Brayne, C., Fox, C., & Boustani, M. (2007). Dementia screening in primary care: Is it time?. *Journal of the American Medical Association, 298*(20), 2409-2411.

Brodaty, H., Low, L. F., Gibson, L., & Burns, K. (2006). What is the best dementia screening instrument for general practitioners to use?. *The American Journal of Geriatric Psychiatry, 14*(5), 391-400.

Buschke, H., Kuslansky, G., Katz, M., Stewart, W. F., Sliwinski, M. J., Eckholdt,

H. M., & Lipton, R. B. (1999). Screening for dementia with the memory impairment screen. *Neurology*, *52*(2), 231-231.

Commenges, D., Gagnon, M., Letenneur, L., Dartigues, J. F., Barberger-Gateau, P., & Salamon, R. (1992). Improving Screening for Dementia in the Elderly Using Mini-Mental State Examination Subscores, Benton's Visual Retention Test, and Isaacs' Set Test. *Epidemiology*, *3*(2), 185-188.

Cullum, C. M. & Lacritz, L. H. (2009). Neuropsychological Assessment in Dementia. In M. F. Weiner & A. M. Lipton (Eds.), *Alzheimer disease and other dementias* (85-104). American Psychiatric Pub.

Cullum, C.M., Thompson, L.L., & Smernoff, E.N. (1993). Three word recall as a measure of memory. *Journal of Clinical and Experimental Neuropsychology, 15*(2), 321-329.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). California Verbal Learning Test – Second Edition. Adult Version. Manual. *Psychological Corporation*.

Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., ... & Aisen, P. S. (2014). The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurology*, *71*(8), 961-970.

Dudas, R. B., Clague, F., Thompson, S. A., Graham, K. S., & Hodges, J. R.

(2005). Episodic and semantic memory in mild cognitive impairment. *Neuropsychologia*, *43*(9), 1266-1276.

Epker, M. O., Lacritz, L. H., & Cullum, C. M. (1999). Comparative analysis of qualitative verbal fluency performance in normal elderly and demented populations. *Journal of Clinical and Experimental Neuropsychology*, *21*(4), 425-434.

Faustman, W. O., Moses, J. A., & Csernansky, J. G. (1990). Limitations of the Mini-Mental State Examination in predicting neuropsychological functioning in a psychiatric sample. *Acta Psychiatrica Scandinavica*, *81*(2), 126-131.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189-198.

Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2013). Montreal cognitive assessment: validation study for mild cognitive impairment and Alzheimer disease. *Alzheimer Disease & Associated Disorders*, *27*(1), 37-43.

Freitas, S., Simões, M. R., Alves, L., Vicente, M., & Santana, I. (2012). Montreal Cognitive Assessment (MoCA): validation study for vascular dementia. *Journal of the International Neuropsychological Society*, *18*(6), 1031-1040.

Fujiwara, Y., Suzuki, H., Yasunaga, M., Sugiyama, M., Ijuin, M., Sakuma, N., ...

& Shinkai, S. (2010). Brief screening tool for mild cognitive impairment in older Japanese: validation of the Japanese version of the Montreal Cognitive Assessment. *Geriatrics & Gerontology International*, *10*(3), 225-232.

Galasko, D., Klauber, M. R., Hofstetter, C. R., Salmon, D. P., Lasker, B., & Thal, L. J. (1990). The Mini-Mental State Examination in the early diagnosis of Alzheimer's disease. *Archives of Neurology*, *47*(1), 49-52.

Hachinski, V., Iadecola, C., Petersen, R. C., Breteler, M. M., Nyenhuis, D. L., Black, S. E., … Leblanc, G. G. (2006). National Institute of Neurological Disorders and Stroke–Canadian stroke network vascular cognitive impairment harmonization standards. *Stroke*, *37*(9), 2220-2241.

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

Haubois, G., Annweiler, C., Launay, C., Fantino, B., de Decker, L., Allali, G., & Beauchet, O. (2011). Development of a short form of Mini-Mental State Examination for the screening of dementia in older adults with a memory complaint: a case control study. *BMC Geriatrics*, *11*, 59.

Heaton, R., Miller, W., Taylor, M., & Grant, I. (2004). Revised comprehensive norms for an expanded Halstead-Reitan battery: demographically adjusted

neuropsychological norms for African-American and Caucasian adults. Lutz, FL: Psychological Assessment Resources, Inc.

Iliffe, S., Robinson, L., Brayne, C., Goodman, C., Rait, G., Manthorpe, J., & Ashley, P. (2009). Primary care and dementia: 1. diagnosis, screening and disclosure. *International Journal of Geriatric Psychiatry*, *24*(9), 895-901.

Ismail, Z., Rajji, T. K., & Shulman, K. I. (2010). Brief cognitive screening instruments: an update. *International Journal of Geriatric Psychiatry*, *25*(2), 111-120.

Kalbe, E., Kessler, J., Calabrese, P., Smith, R., Passmore, A. P., Brand, M. A., & Bullock, R. (2004). DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *International Journal of Geriatric Psychiatry*, *19*(2), 136-143.

Karrasch, M., Sinervä, E., Grönholm, P., Rinne, J., & Laine, M. (2005). CERAD test performances in amnestic mild cognitive impairment and Alzheimer's disease. *Acta Neurologica Scandinavica*, *111*(3), 172-179.

Kinsella, K. & Wan, H. U.S. Census Bureau: International Population Reports, P95/09-1, An Aging World: 2008. US Census Bureau. Washington, DC: U.S. Government Printing Office; 2009.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, *14*(2), 1137-1145.

Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., Letenneur, L., & Dartigues, J. F. (2000). A cognitive screening battery for dementia in the elderly. *Journal of Clinical Epidemiology*, *53*(10), 980-987.

Landi, F., Tua, E., Onder, G., Carrara, B., Sgadari, A., Rinaldi, C., ... & Bernabei, R. (2000). Minimum data set for home care: a valid instrument to assess frail older people living in the community. *Medical Care*, *38*(12), 1184-1190.

Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Randolph, C. (1999). An empirically derived short form of the Boston naming test. *Archives of Clinical Neuropsychology*, *14*(6), 481-487.

Larner, A. J. (2007). DemTect: 1-year experience of a neuropsychological screening test for dementia. *Age and Ageing*, *36*(3), 326-327.

Lee, J. Y., Lee, D. W., Cho, S. J., Na, D. L., Jeon, H. J., Kim, S. K., ... & Cho, M. J. (2008). Brief screening for mild cognitive impairment in elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology*, *21*(2), 104-110.

Levy, J., & Heller, W. (1992). Gender differences in human neuropsychological

function. In A. A. Gerall, H. Moltz, & I. Ward (Eds.), *Sexual differentiation: Handbook of behavioral neurobiology* (Vol. 11, pp. 245-274). New York: Plenum Press.

Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. *Neuropsychology, 15*(2), 165-173.

Lonie, J. A., Tierney, K. M., & Ebmeier, K. P. (2009). Screening for mild cognitive impairment: a systematic review. *International Journal of Geriatric Psychiatry, 24*(9), 902-915.

Lorentz, W. J., Scanlan, J. M., & Borson, S. (2002). Brief screening tests for dementia. *Canadian Journal of Psychiatry, 47*(8), 723-733.

Markwick, A., Zamboni, G., & de Jager, C. A. (2012). Profiles of cognitive subtest impairment in the Montreal Cognitive Assessment (MoCA) in a research cohort with normal Mini-Mental State Examination (MMSE) scores. *Journal of Clinical and Experimental Neuropsychology, 34*(7), 750-757.

Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., & Hodges, J. R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology, 55*(11), 1613-1620.

McDowell, I., Kristjansson, B., Hill, G. B., & Hebert, R. (1997). Community

screening for dementia: the Mini Mental State Exam (MMSE) and modified Mini-mental State Exam (3MS) compared. *Journal of Clinical Epidemiology*, *50*(4), 377-383.

Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of psychiatric research*, *43*(4), 411-431.

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, *21*(11), 1078-1085.

Monsch, A. U., Bondi, M. W., Butters, N., Paulsen, J. S., Salmon, D. P., Brugger, P., & Swenson, M. R. (1994). A comparison of category and letter fluency in Alzheimer's disease and Huntington's disease. *Neuropsychology*, *8*(1), 25.

Monsch, A. U., Bondi, M. W., Butters, N., Salmon, D. P., Katzman, R., & Thal, L. J. (1992). Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Archives of Neurology*, *49*(12), 1253-1258.

Morris, J. C., Mohs, R. C., Rogers, H., Fillenbaum, G., & Heyman, A. (1988). CERAD clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacology Bulletin*, *24*, 641-651.

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

National Institute for Health and Clinical Excellence and Social Care Institute for Excellence (2006). Dementia: Supporting People with Dementia and their Carers. Clinical Practice Guidelines. London: NICE.

Ott, A., Breteler, M. M., van Harskamp, F., Stijnen, T., & Hofman, A. (1998). Incidence and risk of dementia: the Rotterdam Study. *American Journal of Epidemiology*, *147*(6), 574-580.

Pencina, M. J., D'Agostino, R. B., D'Agostino, R.B., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, *27*(2), 157-172.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183-194.

Pinto, E., & Peters, R. (2008). Literature review of the Clock Drawing Test as a tool for cognitive screening. *Dementia and Geriatric Cognitive Disorders*, *27*(3), 201-213.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013).

The global prevalence of dementia: a systematic review and meta-analysis. *Alzheimer's & Dementia*, *9*(1), 63-75.

Rahman, T. T. A., Gaafary, E., & Mohamed, M. (2009). Montreal Cognitive Assessment Arabic version: reliability and validity prevalence of mild cognitive impairment among elderly attending geriatric clubs in Cairo. *Geriatrics & Gerontology International*, *9*(1), 54-61.

Ravaglia, G., Forti, P., Maioli, F., Servadei, L., Martelli, M., Brunetti, N., ... & Mariani, E. (2005). Screening for mild cognitive impairment in elderly ambulatory patients with cognitive complaints. *Aging Clinical and Experimental Research*, *17*(5), 374-379.

Ribeiro, F., Guerreiro, M., & De Mendonça, A. (2007). Verbal learning and memory deficits in Mild Cognitive Impairment. *Journal of Clinical and Experimental Neuropsychology*, *29*(2), 187-197.

Riffenburgh, R. H. (2006). Statistics in medicine. New York, NY: Elsevier Inc

Roalf, D. R., Moberg, P. J., Xie, S. X., Wolk, D. A., Moelter, S. T., & Arnold, S. E. (2013). Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. *Alzheimer's & Dementia*, *9*(5), 529-537.

Roivainen, E. (2011). Gender differences in processing speed: A review of recent research. *Learning and Individual Differences*, *21*(2), 145-149.

Rossetti, H. C., Lacritz, L. H., Cullum, C. M., & Weiner, M. F. (2011). Normative

data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology, 77*(13), 1272-1275.

Salmon, D. P. (2012). Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. *Behavioral Neurobiology of Aging* (pp. 187-212). Springer Berlin Heidelberg.

Salmon, D. P., & Lange, K. L. (2001). Cognitive screening and neuropsychological assessment in early Alzheimer's disease. *Clinics in Geriatric Medicine, 17*(2), 229-254.

Schultz-Larsen, K., Kreiner, S., & Lomholt, R. K. (2007). Mini-Mental Status Examination: mixed Rasch model item analysis derived two different cognitive dimensions of the MMSE. *Journal of Clinical Epidemiology, 60*(3), 268-279.

Schultz-Larsen, K., Lomholt, R. K., & Kreiner, S. (2007). Mini-Mental Status Examination: a short form of MMSE was as accurate as the original MMSE in predicting dementia. *Journal of Clinical Epidemiology, 60*(3), 260-267.

Shulman, K. I. (2000). Clock-drawing: is it the ideal cognitive screening test?.*International Journal of Geriatric Psychiatry, 15*(6), 548-561.

Shulman, K. I., Herrmann, N., Brodaty, H., Chiu, H., Lawlor, B., Ritchie, K., &

Scanlan, J. M. (2006). IPA survey of brief cognitive screening instruments. *International Psychogeriatrics*, *18*(02), 281-294.

Smith, T., Gildeh, N., & Holmes, C. (2007). The Montreal Cognitive Assessment: validity and utility in a memory clinic setting. *Canadian Journal of Psychiatry*, *52*(5), 329.

Spreen, O., & Benton, A. L. (1977). *Neurosensory center comprehensive examination for aphasia* (p. 1969). University of Victoria; Neuropsychology Laboratory.

Strub, R. L., & Black, F. W. (1985) The mental status examination in neurology. Philidelphia, PA: FA Davis Company.

Tabert, M. H., Manly, J. J., Liu, X., Pelton, G. H., Rosenblum, S., Jacobs, M., ... & Devanand, D. P. (2006). Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Archives of General Psychiatry*, *63*(8), 916-924.

Tangalos, E. G., Smith, G. E., Ivnik, R. J., Petersen, R. C., Kokmen, E., Kurland, L. T., … Parisi, J. E. (1996). The Mini-Mental State Examination in general medical practice: clinical utility and acceptance. *Mayo Clinic Proceedings*, *71*(9), 829-837.

Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *The Journal of Clinical Psychiatry*, *48*(8), 314.

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: a

comprehensive review. *Journal of the American Geriatrics Society (40)*9, 922-935.

Twamley, E. W., & Bondi, M. W. (2004). The Differential Diagnosis of Dementia. In J. H. Ricker (Ed.), *Differential diagnosis in adult neuropsychological assessment* (276-326). Springer Publishing Company.

Uhlmann, R. F., & Larson, E. B. (1991). Effect of education on the mini-mental state examination as a screening test for dementia. *Journal of the American Geriatrics Society, 39*(9), 876-880.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin, 83*(2), 213.

Wang, Q. S., & Zhou, J. N. (2002). Retrieval and encoding of episodic memory in normal aging and patients with mild cognitive impairment. *Brain Research*, *924*(1), 113-115.

Wechsler, D. (1997). *WAIS-III: Administration and scoring manual: Wechsler adult intelligence scale*. Psychological Corporation.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*.

Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, *2*(4), a006171.

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., &

Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences*, *35*(4), 863-875.

Welsh, K., Butters, N., Hughes, J., Mohs, R., & Heyman, A. (1991). Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Archives of Neurology, 48*(3), 278-281.

Wind, A. W., Schellevis, F., G., Van Staveren, G. E., Scholten, R. J., Jonker, C., & Van Eijk, J. T. (1997). Limitations of the Mini-Mental State Examination in diagnosing dementia in general practice. *International Journal of Geriatric Psychiatry, 12*(1), 101-108.

Xu, G., Meyer, J. S., Thornby, J., Chowdhury, M., & Quach, M. (2002). Screening for mild cognitive impairment (MCI) utilizing combined mini-mental-cognitive capacity examinations for identifying dementia prodromes. *International Journal of Geriatric Psychiatry*, *17*(11), 1027-1033.

Zakzanis, K. K., Leach, L., & Kaplan, E. (1999). *Neuropsychological differential diagnosis*. Swets & Zeitlinger Publishers.