Laptop-Administered NIH Toolbox and Cogstate Brief Battery in Community-Dwelling Black Adults: Unexpected Pattern of Cognitive Performance between MCI and Healthy Controls

Anson Kairys[1],* , Ana Daugherty[2,3], Voyko Kavcic[2], Sarah Shair[4], Carol Persad[1], Judith Heidebrink[4], Arijit Bhaumik[1] and Bruno Giordani[1,4]

[1]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA
[2]Institute of Gerontology, Wayne State University, Detroit, MI, USA
[3]Department of Psychology, Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI, USA
[4]Department of Neurology, University of Michigan, Ann Arbor, MI, USA

Abstract

Objective: Black adults are approximately twice as likely to develop Alzheimer's disease (AD) than non-Hispanic Whites and access diagnostic services later in their illness. This dictates the need to develop assessments that are cost-effective, easily administered, and sensitive to preclinical stages of AD, such as mild cognitive impairment (MCI). Two computerized cognitive batteries, NIH Toolbox-Cognition and Cogstate Brief Battery, have been developed. However, utility of these measures for clinical characterization remains only partially determined. We sought to determine the convergent validity of these computerized measures in relation to consensus diagnosis in a sample of MCI and healthy controls (HC). **Method:** Participants were community-dwelling Black adults who completed the neuropsychological battery and other Uniform Data Set (UDS) forms from the AD centers program for consensus diagnosis (HC = 61; MCI = 43) and the NIH Toolbox-Cognition and Cogstate batteries. Discriminant function analysis was used to determine which cognitive tests best differentiated the groups. **Results:** NIH Toolbox crystallized measures, Oral Reading and Picture Vocabulary, were the most sensitive in identifying MCI apart from HC. Secondarily, deficits in memory and executive subtests were also predictive. UDS neuropsychological test analyses showed the expected pattern of memory and executive functioning tests differentiating MCI from HC. **Conclusions:** Contrary to expectation, NIH Toolbox crystallized abilities appeared preferentially sensitive to diagnostic group differences. This study highlights the importance of further research into the validity and clinical utility of computerized neuropsychological tests within ethnic minority populations.

Keywords: Memory, Computers, Cognition, Dementia, Ethnic groups, Diagnosis

INTRODUCTION

With the significant growth in the population aged 65 years and older, the number of Americans with Alzheimer's disease (AD) and other dementias is projected to increase from 58 million in 2018 to 88 million by 2050 (Alzheimer's Association, 2018). As these figures grow, disparities have emerged in the prevalence of AD. As compared to non-Hispanic Whites, Black adults have higher rates of dementia per capita and are approximately twice as likely to have AD or other dementias (Gurland et al., 1999; Potter et al., 2009).

There are also considerable barriers to treatment; Black adults are diagnosed later in the disease (Chin, Negash, & Hamilton, 2011) and are less likely to receive antidementia medications (Zuckerman et al., 2008) as compared to non-Hispanic Whites. This evident disparity that places Black adults at higher risk for developing AD and barriers to treatment has been attributed to social and behavioral determinants of health: perceived discrimination and environmental stress (Zahodne, Sol, & Kraal, 2019); a higher incidence of comorbid modifiable health risks (i.e., hypertension, cardiovascular disease, diabetes; Brancati et al., 2000; Sundquist, Winkleby, & Pudaric, 2001; Cushman et al., 2008); and insensitive assessment related to years of education, educational quality, and literacy (Manly, Schupf, Tang, & Stern, 2005; Manly, Touradji, Tang, & Stern, 2003).

*Correspondence and reprint requests to: Anson Kairys, PhD, Neuropsychology Section, Department of Psychiatry, University of Michigan, 2101 Commonwealth Blvd., Suite C., Ann Arbor, MI 48105, USA. Email: aekairys@med.umich.edu

Due to the increased prevalence of dementia and importance of early detection for possible intervention, there remains a critical need for assessments that are sensitive to preclinical signs of AD, especially among persons at high risk. Towards this goal, assessments that are sensitive to identify mild cognitive impairment (MCI) may be useful for clinical diagnosis and research in the development of interventions to promote cognitive resiliency. MCI is often a preclinical stage of AD, and individuals with MCI are more likely to develop AD or other dementias (Petersen et al., 2018). MCI clinical diagnostic criteria include cognitive impairments that are not normal for the individual's age, but not yet severe enough to cause significant impairment in instrumental activities of daily living (Winblad et al., 2004). MCI has been divided into subtypes based on the number of affected cognitive domains and whether memory is affected, that is, single or multidomain amnestic (aMCI) and nonamnestic (naMCI) subtypes. An estimated 15.8% of adults aged 60 and older who are residing in the USA are diagnosed with some subtype of MCI (Petersen et al., 2018), and there is evidence suggesting a higher prevalence of naMCI among Black adults as compared to non-Hispanic Whites (Lopez, Jagust, DeKosky et al., 2003; Katz et al., 2012).

Despite the increased risk posed to Black adults for developing AD, Black adults are largely underrepresented in the research seeking to understand or treat these conditions. Out of 10 of the most prominent biomarker studies on AD in the USA, half did not achieve adequate samples of Black adults required to determine racial/ethnic differences effectively (Shin & Doraiswamy, 2016). It has been proposed that community-based approaches to recruitment which focus on building trust in the research process involve, for example, researchers of the same ethnicity as the participants and use visit locations and timing that are convenient for the participants to reduce the burden of participation (Gilmore-Bykovskyi et al., 2019). These methods have been proposed as ways to help reduce both selection and sampling bias when attempting to evaluate differing rates of disease progression across racial/ethnic groups in large AD center datasets (Gleason et al., 2019). Community-based assessment can also be facilitated by computer administration of cognitive tests due to the ease of accessibility, transportability, and the reduced amount of time and effort required to complete such assessments (Hinton et al., 2010). Limited research suggests that older Black adults are more highly satisfied with computerized neuropsychological assessment than paper–pencil methods and that individual levels of familiarity and comfort with computers do not affect performance on assessments of cognition (Gamaldo et al., 2018). While this highlights the strengths of a computerized assessment that can be used in the community, the validity and sensitivity of these methods to identify individuals with MCI- or AD-related cognitive symptoms, specifically among Black adults, have not been thoroughly evaluated.

Two computerized cognitive assessment batteries that are increasingly being used in clinical research are the Cogstate Brief Battery (CBB; Hammers et al., 2012; Hammers et al., 2011; Maruff et al., 2013) and the NIH Toolbox-Cognition Battery (NIHTB-CB; Heaton et al., 2014; Weintraub et al., 2013). These measures have demonstrated acceptable reliability and construct validity as compared to traditional paper–pencil methods (Casaletto et al., 2015; Heaton et al., 2014) but lack clear support as a replacement for gold-standard neuropsychological assessment (Scott, Sorrell, & Benitez, 2019). Although there has been significant effort put forth to develop representative normative samples for these measures, there continues to be a lack of knowledge around the performance on these two cognitive batteries by individuals from underrepresented populations. For example, two of the largest studies on the psychometric properties of the CBB make no mention of the race/ethnicity of their samples (Maruff et al., 2013; Mielke et al., 2015), pointing to a lack of information regarding the use of these measures in ethnically diverse populations. With regard to the NIHTB-CB, the original normative data were unadjusted for demographic factors, with subsequent normative samples providing corrections for race/ethnicity (Casaletto et al., 2015); however, it has been found that the NIHTB-CB may overestimate such individuals' deficits (Scott et al., 2019) and more research is needed to validate the clinical utility of this measure. Overall, there remains a fundamental gap in our understanding of the utility of these computerized cognitive batteries for use in clinical research trials including Black adults.

The present study was designed to evaluate the convergent validity of the CBB and NIHTB-CB with the current gold-standard consensus diagnosis process from the National Alzheimer's Coordinating Center (NACC) and the Uniform Data Set (UDS). The aim of the study was to determine which individual subtests from these computerized cognitive measures would reflect the greatest difference between MCI and healthy controls (HC) in a sample of community-dwelling Black adults. Based on typical cognitive symptoms of AD and related dementia, we hypothesized that measures of fluid ability (i.e., declarative memory, executive functioning, attention, and processing speed) from the CBB and NIHTB-CB would most strongly differentiate between diagnosis groups as compared to measures of crystallized ability (e.g., vocabulary).

METHOD

Participants

All participants consented to participate in the research study, and procedures were approved by the University of Michigan Medical School Review Board (IRBMED) and/or Wayne State University Research Subjects Review Board. Participants were 113 community-dwelling Black adults aged 56–90 years with self-reported cognitive complaints and no previous neurocognitive diagnosis. Participants were enrolled based on their responses to a question asking if they had experienced "a change in memory or other cognitive areas over the past year, but not so severe as to interfere with ability to complete

Table 1. Sample description

Variable	HC ($n = 61$)	MCI ($n = 43$)	Group comparison (*t*-test, χ^2)
Age (years)	70.97 (6.76)	73.47 (7.15)	−1.81
Sex (female, %)	90.20	86.00	0.42
Education (years)	15.33 (2.34)	14.35 (2.42)	2.07*
Computer anxiety	39.98 (15.73)	42.55 (16.24)	−0.80

HC = healthy controls; MCI = combined group amnestic and nonamnestic mild cognitive impairment.
Note: Descriptive statistics are reported for each group as *M* (*SD*) and percentage of participants who self-identify as female. Statistical comparisons between groups are reported at *t*-tests or chi-square statistics, with significance testing * $p < .05$.

daily activities." Recruitment was conducted through the Wayne State University, Michigan Center for Urban African American Aging Research—Healthier Black Elders Center, and the Michigan Alzheimer's Disease Research Center (Michigan ADRC). All participants completed the NACC-UDS evaluation, a multidomain medical, neurological, social, and neuropsychological evaluation used as the basis for consensus conference diagnostic procedures (Rahman-Filipiak, Giordani, Heidebrink, Bhaumik, & Hampstead, 2018; Weintraub et al., 2018; Weintraub et al., 2009). Participants also completed the NIHTB-CB and CBB within 3 months of their NACC-UDS assessment, the results of which were not used as part of the consensus conference diagnostic process. Patients were excluded if they suffered from a significant psychiatric, medical, or neurological deficit other than AD that may impair cognitive ability. Those with physical limitations which precluded completion of neuropsychological measures were excluded. Nine participants were excluded from the analyses as they did not meet MCI diagnostic criteria based on NACC-UDS criteria but were judged to be cognitively impaired (impaired but not MCI). These criteria parallel the MCI criteria published by the 2011 National Institute of Health and Alzheimer's Association workgroup on MCI and AD (Albert et al., 2011). The final sample for analysis included 61 HC and 43 MCI patients ($n = 29$ aMCI and $n = 14$ naMCI), which were majority multiple domains (Table 1).

Assessment Measures

NACC-UDS Neuropsychological Battery

The NACC-UDS neuropsychological battery was initially developed in 2005 and recently underwent a third revision (Weintraub et al., 2018). The NACC-UDS has been implemented by the National Institute on Aging Alzheimer Disease Centers nationally since its inception, and has been described extensively before (Weintraub et al., 2009). Briefly, the neuropsychological battery, which is used for case-consensus diagnosis, includes measures of dementia severity, learning and memory, vocabulary and reading skills,

verbal fluency, processing speed, executive functioning, and visuospatial perception and memory. In the present study, we included data from the NACC-UDS version two, with supplemental data including the Wide Range Achievement Test 4th edition, reading subtest (WRAT-IV; Wilkinson & Robertson, 2006) and the Wisconsin Card Sorting Test (WCST; Heaton, 1993).

NIH Toolbox-Cognition Battery

The NIHTB-CB is a multidimensional measure comprised of seven subtests and provides individual subtest performances as well as composite summary scores of crystallized cognitive abilities, fluid cognition, and total cognition. The crystallized cognition composite includes the Oral Reading Recognition (ORR) and Picture Vocabulary (PV) subtests. Measures of fluid abilities include the Dimensional Change Card Sort (DCCS) task, Flanker Inhibitory Control and Attention (FICA), List Sorting Working Memory (LSWM), Pattern Comparison Processing Speed (PCPS), and Picture Sequence Memory (PSM) subtests. The specific test details, procedures, and extensive psychometric evaluation are available elsewhere (Weintraub et al., 2013). The DCCS test assesses set shifting and requires participants to match a target stimulus to one of two choice stimuli according to shape and color. The FICA task tests an individual's ability to inhibit visual attention to task-irrelevant dimensions. The LSWM test involves individual presentation of a series of stimuli and requires the participant to repeat them in order from smallest to biggest by category and size. The PSM test is a measure of episodic memory based on ordering pictures in sequence. The ORR test measures participant's ability to pronounce single printed words out loud, and recognize letters. The PV test assesses language by asking participants to pick a picture that matches a spoken word. Finally, the PCPS test requires participants to identify whether two visual patterns are the same or different. In our dataset, all resultant subtest scores were fully adjusted for age, sex, and education. Summary scores were not included in the analyses.

Cogstate Brief Battery

The CBB is a computerized cognitive assessment that provides measures of four different cognitive domains using playing card paradigms: visual learning, working memory, processing speed, and attention. These separate tests and their psychometric properties have been described previously (Falleti, Maruff, Collie, & Darby, 2006; Lim et al., 2013; Maruff et al., 2013). Briefly, the core tests include the Detection Task (DET), a simple reaction time task, the Identification Task (IDN) which is a choice reaction time test of visual attention, the One Card Learning Task (OCL), a continuous visual recognition learning task, and the One Back Task (ONB), a test of working memory. We included data on both accuracy (correct *vs*. incorrect responses) and reaction time (in milliseconds) in our analyses.

Computer Anxiety

Computer anxiety was measured using the Wild et al. (2012) Computer Anxiety Survey, a 16-item measure on which participants rate their level of anxiety when using computers (e.g., "I feel relaxed when I am working on a computer"). Responses are rated on a five-point, Likert-type scale and range from "Strongly Disagree" to "Strongly Agree." Total scores range from 16 to 80, with higher scores indicating greater levels of computer anxiety. Computer anxiety summary scores were derived by totaling rating for each item (Wild et al., 2012).

Statistical Analysis

Hypotheses were tested with a discriminant function analysis (DFA). DFA is a data-driven approach that benefits from *a priori* selection of variables to describe multivariate differences between groups, but is agnostic to the relative importance of each variable. A discriminant function is a composite that describes observed between-group variance by a set of predictor variables. The procedure estimates a number of discriminant functions equal to the minimum degrees of freedom; combined, the functions account for all of the between-group variance and the first function accounts for the largest proportion. In analysis of two groups, a single function accounts for all of the between-group variance. A significant model indicates a reliable difference between groups, and the proportional variance explained by each function is further evaluated. Each predictor variable has a standardized loading in the structure matrix, that, is its unique contribution to the discriminant function; absolute values at least 0.30 are considered meaningful by convention (i.e., at least 9% commonality with the other variables to differentiate the groups). Separation of the group centroids on a discriminant function is consistent with between-group differences in that multivariate combination of data. Applying cross-validation procedures, group membership can be predicted from the loadings of each variable across functions, which is used to assess model accuracy. We apply this technique to describe the differences between HC and MCI in terms of performance on NIHTB-CB and CBB tests and to evaluate the unique contribution of each measure in rank order.

Prior to analysis, all measures were screened for univariate and multivariate outliers, and multivariate normality was assessed with visual inspection of Q–Q plots. One case was identified as a statistical outlier on DCCS and DET tasks and no other univariate outliers were observed; as there was no evidence of multivariate outliers and the assumption of normality was reasonably met, this case was included in analysis. In each DFA analysis, all variables were entered in a single step and included control variables: age, sex, years of education, and computer anxiety summary score. Assumed prior probabilities for group classification were computed from frequency of diagnosis observed in the sample. Of the eligible sample, nine participants were missing data from one or more subtests (e.g. computer malfunction) and eight participants fell below recommended CBB minimum accuracy percentage on one subtest (Cogstate, 2020). Data were missing at random (Little's χ^2 (730, $N = 104$) = 257.18, $p = .99$); those missing data values were replaced by the sample mean in the DFA procedure.

A preliminary analysis applied DFA to confirm diagnosis that was made by consensus conference of the NACC-UDS test battery. This analysis serves as a demonstration of the method and to validate the approach to test between-group differences with the exploratory technique. In our primary hypothesis test, measures of the NIHTB-CB and CBB were tested as predictors in the DFA that were expected to replicate the pattern of results with the NACC-UDS in the preliminary analysis. Namely, performance on tests of fluid cognitive ability would identify participants with MCI apart from HC. Although combined in the primary analysis, aMCI and naMCI are prodromal stages of different dementias, and therefore we also conducted a supplementary DFA analysis to determine if performances may depend on MCI subtype. Models were assessed for significance of the discriminant function and accuracy of classification with cross-validation. Unique contributions of each variable, controlling for all other variables in the model, were interpreted from rank-order structure matrix coefficients.

RESULTS

Participants diagnosed with aMCI ($n = 29$) performed equivalent to those with naMCI ($n = 14$) on all CBB ($t = -1.27$–1.38, all p's $\geq .10$) and NIHTB-CB tests ($t = -1.91$–2.97, all p's $\geq .06$), except the Flanker subtest (t (39) = 2.97, $p = .01$). Due to the low incidence of naMCI observed in this sample and the statistical equivalence to aMCI, primary analyses tested hypotheses with a combined group of aMCI and naMCI, described hereafter as the MCI group, as compared to HC counterparts.

Confirmation of MCI Diagnosis from the UDS

A preliminary analysis applied the DFA method to confirm the diagnosis made by consensus conference when reviewing performance on the NACC-UDS (version 2) neuropsychological battery and ancillary tests (WRAT-IV and WCST). The collective performance on the test battery and control variables significantly described differences between MCI and HC participants: Wilk's $\lambda = 0.27$, χ^2 (27, $N = 104$) = 55.17, $p = .001$. Reviewing the structure matrix, discriminating variables included delayed story recall (0.52), immediate story recall (0.40), digit-symbol coding, a measure of processing speed (0.35), and a test of oral word reading (WRAT-IV; 0.30). All other variables made a small independent contribution that fell below the 0.30 threshold. The collection of variables correctly classified 78.8% (63.5% cross-validated) of participants, and cross-validated accuracy was similar for HC (63.9%) and MCI (62.8%). Taken together, performance on declarative memory and executive
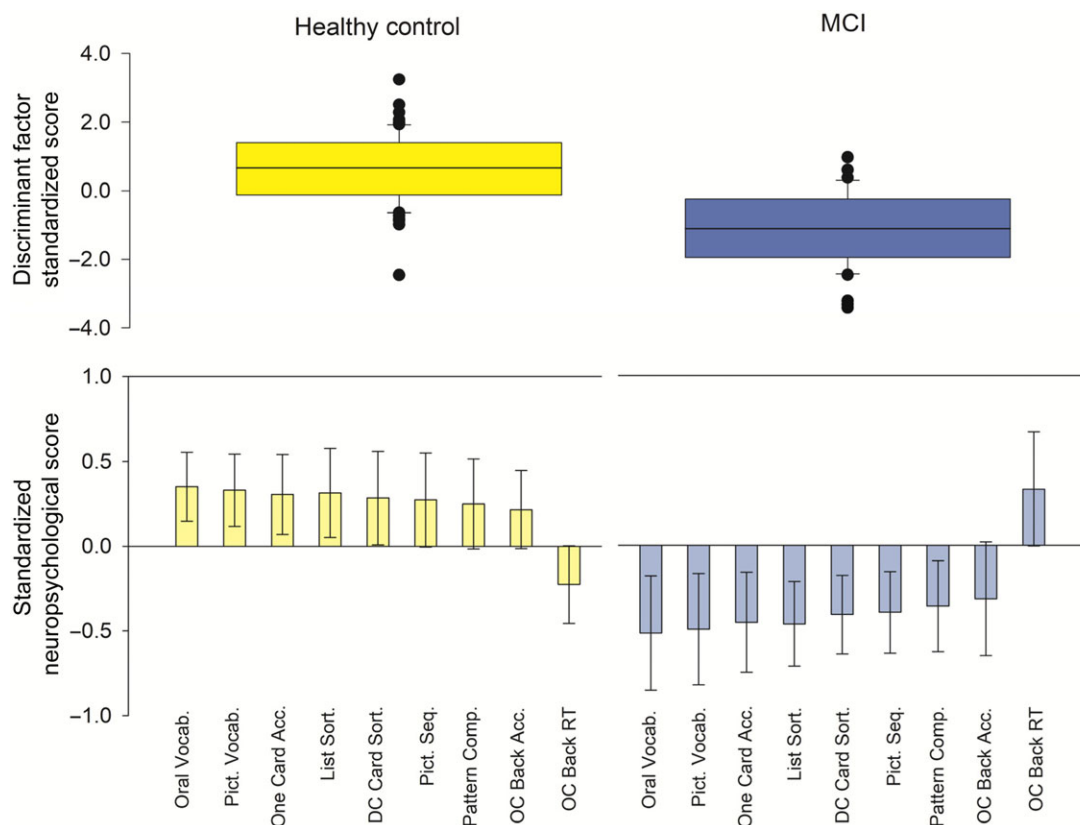
Fig. 1. Multivariate description of performance differences between MCI and HC. Top panel: The discriminant function factor score includes all variables submitted to the DFA. The distribution of scores is shown for HC (yellow) and MCI (blue) groups, including group means (line) and distribution. The model significantly differentiated between groups: Wilk's $\lambda = 0.56$, χ^2 (19, $N = 104$) $= 43.28$, $p = .001$. Bottom panel: The variables that made the strongest, unique contribution to the discriminant function (structure matrix loading $> |0.30|$), are displayed for each group, including error bars that represent 2 standard errors of the group mean. For the purpose of data visualization, scores were standardized to the sample mean. In rank order, NIHTB-CB Oral Vocabulary/Reading and Picture Vocabulary subtests most strongly differentiated between groups, followed by several tests of declarative memory, learning and executive function. Oral Vocab = NIHTB-CB Oral Reading Recognition score; Pict. Vocab. = NIHTB-CB Picture Vocabulary; One Card Acc. = CBB one card learning accuracy; DC Card Sort = NIHTB-CB dimensional change card sort; Pict. Seq. = NIHTB-CB picture sequence; Pattern Comp. = NIHTB-CB pattern comparison processing; OC Back Acc. = CBB one card learning back accuracy; OC Back RT = CBB one card learning back reaction time. NIHTB-CB scores were fully adjusted for demographic characteristics.

function tasks in the NACC-UDS, controlling for all other variables in the model, most strongly distinguished MCI apart from HC. Notably, low WRAT-IV reading scores differentiated MCI participants independent of age and education, but it was of lower rank order importance as compared to the other cognitive measures.

Evaluation of Performance on NIHTB-CB and CBB

Repeating the analysis with the NIHTB-CB and CBB batteries, which were not used to make diagnoses, revealed a different pattern of group differences in performance. The model significantly differentiated between groups: Wilk's $\lambda = 0.56$, χ^2 (19, $N = 104$) $= 43.28$, $p = .001$. In rank order, the strongest discriminating variables were NIHTB-CB Oral Reading (0.55) and NIHTB-CB PV

(0.51). Accounting for these effects, several measures of declarative memory, learning, and executive function further differentiated groups: NIHTB-CB List Sorting (0.41), CBB One Card Learning accuracy (0.41), NIHTB-CB DCCS (0.38), NIHTB-CB Pattern Comparison (0.36), NIHTB-CB PSM (0.34), CBB One Card Back accuracy (0.31), and reaction time (−0.30). The collective performance on these tests indicated participants with MCI performed worse than HC (Figure 1). All other performance measures and control variables contributed small independent effects that fell below threshold (Table 2). The model correctly classified 77.9% (66.3% cross-validated) of cases. The cross-validated accuracy was lower for MCI classification (55.8%) as compared to HC (73.8%).

Our supplementary analysis of performances on the NIHTB-CB and CBB with three groups (aMCI, naMCI, and HC) largely replicated the result pattern. With three

Table 2. Variables that significantly differentiated between MCI and healthy controls

Variable	Discriminant function standardized score
NIHTB-CB Oral Reading Recognition*	0.55
NIHTB-CB Picture Vocabulary*	0.51
NIHTB-CB List Sorting*	0.41
CBB One card learning accuracy (log)	0.41
NIHTB-CB Dimensional Change Card Sort*	0.38
NIHTB-CB Pattern Comparison*	0.35
NIHTB-CB Picture Sequence Memory*	0.34
CBB One Back accuracy (log)	0.31
CBB One Back RT (log)	−0.30
NIHTB-CB Flanker Test *	0.25
Education (years)	0.21
Computer-related anxiety	−0.20
CBB Identification RT (log)	−0.16
CBB Detection RT (log)	−0.15
Age (years)	−0.13
Sex	0.11
CBB Identification Accuracy (log)	0.10
CBB Detection Accuracy (log)	−0.01
CBB One Card Learning RT (log)	0.01

Note: All variables contributed to the significant model. Coefficients are standardized structure matrix scores that identified the discriminant function. Coefficients with an absolute value of at least 0.30 were interpreted; positive coefficients indicate cognitive typical adults scored higher (e.g., greater accuracy) as compared to combined aMCI/naMCI, and negative coefficients indicate lower values (e.g., shorter reaction time).* = scores were fully adjusted for age, sex, and education. NIHTB-CB = NIH Toolbox Cognition Battery; CBB = Cogstate Brief Battery; RT = Reaction time (in milliseconds).

groups, two discriminant functions were estimated (Wilk's $\lambda = 0.39$, χ^2 (38, $N = 104$) = 70.36, $p = .001$), although the second function was not significant beyond the first (Wilk's $\lambda = 0.72$, χ^2 (18, $N = 104$) = 24.38, $p = 0.14$). Function 1 was determined by NIHTB-CB PV (0.54), NIHTB-CB Oral Reading (0.46), CBB One Card Learning accuracy (0.41), NIHTB-CB List Sorting (0.37), NIHTB-CB PSM (0.37), and NIHTB-CB DCCS (0.35), and all other variables fell below threshold. The cross-validated classification accuracy was low (53.8%) and poorly identified aMCI (31.0%) and naMCI (14.3%). The pattern of results underscores that NIHTB-CB and CBB tests of declarative memory and executive function differentiated Black adults with MCI as compared to HC, but regardless if amnestic or nonamnestic, NIHTB-CB PV and Oral Reading tests were the strongest unique predictors of MCI.

DISCUSSION

The present evidence indicates that, even when controlling for education and other demographic factors, computerized measures of crystallized intelligence, namely NIHTB-CB measures of Oral Reading and PV skills, were most sensitive to identifying MCI apart from HC in this sample of Black

adults as compared to other NIHTB-CB and CBB tasks. Second to these unique effects, other deficits in memory and executive function were predictive of MCI, which was consistent with the expected preclinical symptoms; but, surprisingly, these were not the strongest descriptors of group differences. The NACC-UDS outcomes that were used for clinical diagnosis via case consensus produced the expected pattern of differences between the groups, foremost in memory and executive functioning deficits in individuals with MCI. Therefore, the results from the CBB and NIHTB-CB assessments appear to reflect differences in the cognitive domain assessments when administered via laptop computer. In sum, while reading and vocabulary scores across assessment tools were lower in Black adults with MCI as compared to HC, the computerized assessments produced a pattern of results that was not fully consistent with what is typically expected for diagnosis of MCI and did not fit with the pattern of results observed on the NACC-UDS measures.

Differences in measures of reading ability and vocabulary skills have been shown between individuals with MCI and HC previously. For example, individuals with MCI perform worse than HC on the Peabody Picture Vocabulary Test, a paper–pencil test which is similar to the PV task administered in the NIHTB-CB (Jokel, Seixas Lima, Fernandez, & Murphy, 2019). Although such vocabulary skills are thought to generally remain stable across the lifespan and represent estimated premorbid cognitive functioning (i.e., crystallized abilities), there is also evidence that performance on the Peabody Picture Vocabulary Test declines with increasing cognitive deficit, such as when comparing individuals with MCI to those with AD (Snitz, Bieliauskas, Crossland, Basso, & Roper, 2000) and could reflect changes in language functioning. The relation of reading ability and literacy to cognitive performance is more complex. Low reading ability may present a bias in verbal assessments of fluid cognitive abilities, for example, NIHTB-CB scores are adjusted for education level. However, it is unclear if a single adjustment can be applied to all persons, regardless of race and ethnicity. In contrast, high educational attainment and literacy may resemble a form of "cognitive reserve"—protective factors which appear to stave off eventual cognitive decline—which may be more predictive of subsequent declines in memory, executive functioning, and language skills than years of education alone (Manly et al., 2005; Manly et al., 2003; Stern, 2002; Stern, 2006).

The pattern of results we report here suggests that crystallized abilities measured by the NIHTB-CB and CBB are relatively more sensitive than NIHTB-CB and CBB tasks of memory and executive functioning as predictors of MCI in this sample of Black adults. Notably, our sample consisted of individuals with an average of 14 years of education, and all analyses identified years of education as having a negligible, independent effect. In the preliminary analysis of the NACC-UDS supplemental data, WRAT-IV reading scores contributed to differentiating between MCI and HC, which suggests that differences in literacy and educational quality (Dotson, Kitner-Triolo, Evans, & Zonderman,

2009) may be relevant when identifying MCI among Black adults, harkening to previous discussions of educational inequality in the USA (Baker, Johnson, Velli, & Wiley, 1996). However, this alone cannot account for the unexpected pattern of results observed in the computerized test batteries. First, the rank order importance of WRAT-IV reading scores was lower than performance on declarative memory and executive function, as would be expected for identifying persons with MCI. Second, the equivalent measures in the NIHTB-CB (Oral Reading and PV) were found to be the strongest unique effects to identify MCI apart from HC, even after scores were adjusted for education and race/ethnicity following standardized procedures. One plausible interpretation of this result is the examined tests are relatively more sensitive to reading ability, which may present as a source of bias when assessing MCI. Alternatively, it may be the case that these computerized measures are relatively less sensitive to memory ability than the NACC-UDS measures, leading to the observed pattern of findings on those measures. These unexpected differences should be evaluated in future studies.

Another interpretation is the possibility that computerized tests are not wholly equivalent to paper–pencil assessments that were the original basis of diagnostic criteria, perhaps due to differences in difficulty level and additional sources of performance error related to computer use. This issue has been raised by Loring et al., (2019), who found differential predictive ability of the NIHTB-CB PSMT and the Rey Auditory Verbal Learning Test (AVLT), with the PSMT accurately predicting impaired AVLT performance, but poorly predicting combined borderline/impaired performance (Loring et al., 2019). This raises the possibility that there is something inherently different about verbal learning tests administered via paper and those administered on a computer, and the same may hold for tests of crystallized ability. Additionally, the NIHTB-CB and the CBB have now been reformatted for iPad administration and although the iPad NIHTB-CB and computer-based formats have been found to be more or less equivalent in non-Hispanic White samples (Mielke et al., 2015), further comparisons are needed to determine if there is a difference between computer and iPad administrations in Black adult samples. Such a comparison is warranted given the evidence that the CBB iPad version yielded slower performance then computer-based administrations (Stricker et al., 2019). We are currently conducting this analysis at the Michigan Alzheimer's Disease Research Center, but more work including comparisons across different racial/ethnic groups is needed to determine the sensitivity and validity of these assessments in characterizing MCI and AD.

Beyond our finding of group discrimination based foremost on crystallized abilities, computerized measures of declarative memory, learning, and executive functioning also significantly contributed to the group discrimination model, albeit to a relatively lesser extent. While memory impairment is emphasized in studies of preclinical AD, evidence suggests that impairment in cognitive domains other than memory, such as executive functioning and language, may be more accurate predictors of early cognitive decline than memory functioning, particularly in Black adults who are living independently in urban settings (Gamaldo, Allaire, Sims, & Whitfield, 2010). This raises the possibility that the preclinical/MCI symptom profile in Black adults may be different than non-Hispanic White adults, but this has yet to be fully elucidated given little demographic diversity in many clinical trials of MCI, AD, and other dementias. Further, participants in the MCI group more closely represented a multidomain MCI profile, rather than a single-domain aMCI, which was consistent with the NACC-UDS consensus diagnoses. Using the computerized batteries and developing cognitive profiles using a data-driven approach is an important step in determining which diagnostic features appear to be universal and which may differ in Black adults as compared to non-Hispanic Whites. Overall, while these computerized assessments appear to be relatively more sensitive to differences in crystallized ability than memory and executive functioning, the extent to which these measures can be used to identify multidomain MCI in Black adults indicates that computerized measures may be suitable for this type of community-based research, as long as relevant control variables are considered. As of now, these computerized measures do not appear to be a suitable substitute for gold-standard neuropsychological assessment and clinical diagnosis.

This study is not without limitations. First, we did not have an adequate sample size within our groups to compare aMCI and naMCI participants; instead we combined these samples into a single MCI group. This was decided based on our preliminary analysis showing no significant differences on cognitive test performances between those groups, except on the NIHTB-CB Flanker subtest. We verified this by showing that the three-group model also yielded a similar pattern of discriminant tests. Further, it is important to emphasize that our MCI participants were, for the most part, multidomain MCI which helps to explain the presence of tasks of executive functioning, learning, and declarative memory as also contributing to our DFA model. Future research will seek to expand on these findings and include a wider range of cognitive-associated diagnoses (e.g., multi- and single-domain naMCI, aMCI, and AD samples). Secondly, our sample was based on volunteers from the community and was predominantly female, which may limit the generalizability of these findings to broader populations outside of an AD research center setting. Importantly, Sundermann et al. (2016) found that females performed better than males on a test of verbal memory, despite comparable brain hypometabolism, and the relationship between hypometabolism and memory performance was stratified such that females were most different from males at a later disease state (i.e., increased hypometabolism) and differences between sexes were minimal at the mild end of the disease state. Therefore, the authors argue that females may have a "cognitive reserve" advantage over males in the domain of verbal memory which is protective against memory decline

later in the disease state. Future research should include more males to further explore that hypothesis. We report a cross-sectional analysis that cannot evaluate the sensitivity of the cognitive assessment to decline and transition to disease; however, the between-group comparison and validation against the NACC-UDS identified critical directions for future research on computerized assessments. Computer-related anxiety was included as covariate in all analyses, but there also remains a question as to whether computerized assessments may differentially bias performance across cognitive domains, especially assessments of crystallized functions (Scott et al., 2019). A more thorough understanding of how this may contribute to differences in sensitivity to detect MCI among Black adults is warranted.

To our knowledge, this is the first report to use both the NIHTB-CB and the CBB in a sample of community-dwelling Black adults in the assessment of MCI. This research expands our understanding of the use of computerized cognitive batteries and hopefully highlights important areas of needed work in utilizing and interpreting findings from these measures. One conclusion and area of further work is the need for more sensitive and specific methods for evaluating ethnic minority populations. Including aspects such as depression and perceived discrimination as it relates to cognitive decline, as others have (Zahodne et al., 2019), will also help us better elucidate our findings. As we continue the important work of seeking novel ways to enhance our understanding of MCI in underrepresented populations and improve recruitment and retention strategies, computerized cognitive assessments will likely play an important role in this endeavor. Therefore, we must ensure that the data we are obtaining are applicable to these populations of interest, are culturally and racially agnostic, and yield useful clinical research outcomes across diverse samples which can serve as comparable outcomes to more formalized assessment batteries such as the NACC-UDS. We will extend on this work by seeking to determine if these results are replicable and through observing the longitudinal trajectories of these individuals to inform our cross-sectional findings.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S135561772100028X

ACKNOWLEGMENTS

FINANCIAL SUPPORT

CONFLICTS OF INTEREST

None.

REFERENCES

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., ... Phelps, C.H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers & Dementia*, *7*(3), 270–279. doi: 10.1016/j.jalz.2011.03.008

Association, A. s. (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *14*(3), 367–429. doi:https://doi.org/10.1016/j.jalz.2018.02.001

Baker, F.M., Johnson, J.T., Velli, S.A., & Wiley, C. (1996). Congruence between education and reading levels of older persons. *Psychiatric Services*, *47*(2), 194–196.

Brancati, F.L., Kao, W.H.L., Folsom, A.R., Watson, R.L., & Szklo, M. (2000) Incident Type 2 Diabetes Mellitus in African American and White Adults: The Atherosclerosis Risk in Communities Study. *JAMA*, *283*(17), 2253–2259. doi: 10.1001/jama.283.17.2253

Casaletto, K.B., Umlauf, A., Beaumont, J., Gershon, R., Slotkin, J., Akshoomoff, N., & Heaton, R.K. (2015). Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*, *21*(5), 378–391. doi: 10.1017/s1355617715000351

Chin, A.L., Negash, S., & Hamilton, R. (2011). Diversity and disparity in dementia: the impact of ethnoracial differences in Alzheimer disease. *Alzheimer Disease and Associated Disorders*, *25*(3), 187–195. doi: 10.1097/WAD.0b013e318211c6c9

Cogstate. (2020). *Cogstate Research: Guidelines for Analysis*. Australia: CogState Limited.

Cushman, M., Cantrell, R.A., McClure, L.A., Howard, G., Prineas, R.J., Moy, C.S., Temple, E.M., & Howard, V.J. (2008). Estimated 10-year stroke risk by region and race in the United States: geographic and racial differences in stroke risk. *Annals of Neurology*, *64*(5), 507–513. doi: 10.1002/ana.21493. PMID: 19067365.

Dotson, V.M., Kitner-Triolo, M.H., Evans, M.K., & Zonderman, A.B. (2009). Effects of race and socioeconomic status on the relative influence of education and literacy on cognitive functioning. *Journal of the International Neuropsychological Society*, *15*(4), 580–589. doi: 10.1017/s1355617709090821

Falleti, M.G., Maruff, P., Collie, A., & Darby, D.G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1095–1112. doi: 10.1080/13803390500205718

Gamaldo, A.A., Allaire, J.C., Sims, R.C., & Whitfield, K.E. (2010). Assessing mild cognitive impairment among older African

Americans. *International Journal of Geriatric Psychiatry*, 25(7), 748–755. doi: 10.1002/gps.2417

Gamaldo, A.A., Tan, S.C., Sardina, A.L., Henzi, C., Guest, R., Ross, L.A., ... Andel, R.A. (2018). Older Black Adults' Satisfaction and Anxiety Levels After Completing Alternative Versus Traditional Cognitive Batteries. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences.* doi: 10.1093/geronb/gby095

Gilmore-Bykovskyi, A.L., Jin, Y., Gleason, C., Flowers-Benton, S., Block, L.M., Dilworth-Anderson, P., ... Zuelsdorff, M. (2019). Recruitment and retention of underrepresented populations in Alzheimer's disease research: A systematic review. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 5, 751–770. doi:https://doi.org/10.1016/j.trci.2019.09.018

Gleason, C.E., Norton, D., Zuelsdorff, M., Benton, S.F., Wyman, M.F., Nystrom, N., ... Asthana, S. (2019). Association between enrollment factors and incident cognitive impairment in Blacks and Whites: Data from the Alzheimer's Disease Center. *Alzheimers & Dementia*, 15(12), 1533–1545. doi: 10.1016/j.jalz.2019.07.015

Gurland, B.J., Wilder, D.E., Lantigua, R., Stern, Y., Chen, J.M., Killeffer, E.H.P., & Mayeux, R. (1999). Rates of dementia in three ethnoracial groups. *International Journal of Geriatric Psychiatry*, 14(6), 481–493.

Hammers, D., Spurgeon, E., Ryan, K., Persad, C., Barbas, N., Heidebrink, J., ... Giordani, B. (2012). Validity of a brief computerized cognitive screening test in dementia. *Journal of Geriatric Psychiatry and Neurology*, 25(2), 89–99. doi: 10.1177/0891988712447894

Hammers, D., Spurgeon, E., Ryan, K., Persad, C., Heidebrink, J., Barbas, N., ... Giordani, B. (2011). Reliability of repeated cognitive assessment of dementia using a brief computerized battery. *American Journal of Alzheimers Disease and Other Dementia*, 26(4), 326–333. doi: 10.1177/1533317511411907

Heaton, R. (1993). *Wisconsin Card Sorting Test (WCST)*: Revised and expanded. Odessa, Fla. (P.O. Box 998, Odessa 33556) : Psychological Assessment Resources.

Heaton, R.K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., ... Gershon, R. (2014). Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *Journal of the International Neuropsychological Society*, 20(6), 588–598. doi: 10.1017/S1355617714000241

Hinton, L., Carter, K., Reed, B.R., Beckett, L., Lara, E., DeCarli, C., & Mungas, D. (2010). Recruitment of a community-based cohort for research on diversity and risk of dementia. *Alzheimer Disease and Associated Disorders*, 24(3), 234–241. doi: 10.1097/WAD.0b013e3181c1ee01

Jokel, R., Seixas Lima, B., Fernandez, A., & Murphy, K.J. (2019). Language in Amnestic Mild Cognitive Impairment and Dementia of Alzheimer's Type: Quantitatively or Qualitatively Different? *Dementia and Geriatric Cognitive Disorders Extra*, 9(1), 136–151. doi: 10.1159/000496824

Katz, M.J., Lipton, R.B., Hall, C.B., Zimmerman, M.E., Sanders, A.E., Verghese, J., Dickson, D.W., & Derby, C.A. (2012). Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and Alzheimer dementia in blacks and whites: a report from the Einstein Aging Study. *Alzheimer Disease and Associated Disorders*, 26(4), 335–343. https://doi.org/10.1097/WAD.0b013e31823dbcfc

Lim, Y.Y., Jaeger, J., Harrington, K., Ashwood, T., Ellis, K.A., Stöffler, A., ... Maruff, P. (2013). Three-month stability of the CogState brief battery in healthy older adults, mild cognitive

impairment, and Alzheimer's disease: results from the Australian Imaging, Biomarkers, and Lifestyle-rate of change substudy (AIBL-ROCS). *Archives of Clinical Neuropsychology*, 28(4), 320–330. doi: 10.1093/arclin/act021

Lopez, O.L., Jagust, W.J., DeKosky, S.T., Becker, J.T., Fitzpatrick, A., Dulberg, C., ... Kuller, L.H. (2003). Prevalence and Classification of Mild Cognitive Impairment in the Cardiovascular Health Study Cognition Study: Part 1. *Archives of Neurology*, 60(10), 1385–1389. doi: 10.1001/archneur.60.10.1385

Lopez, O.L., Jagust, W.J., Dulberg, C., Becker, J.T., DeKosky, S.T., Fitzpatrick, A., ... Kuller, L.H. (2003). Risk factors for mild cognitive impairment in the Cardiovascular Health Study Cognition Study: part 2. *Archives of Neurology*, 60(10), 1394–1399. doi: 10.1001/archneur.60.10.1394

Loring, D.W., Bowden, S.C., Staikova, E., Bishop, J.A., Drane, D.L., & Goldstein, F.C. (2019). NIH Toolbox Picture Sequence Memory Test for Assessing Clinical Memory Function: Diagnostic Relationship to the Rey Auditory Verbal Learning Test. *Archives of Clinical Neuropsychology*, 34(2), 268–276. doi: 10.1093/arclin/acy028

Manly, J.J., Schupf, N., Tang, M.-X., & Stern, Y. (2005). Cognitive Decline and Literacy Among Ethnically Diverse Elders. *Journal of Geriatric Psychiatry and Neurology*, 18(4), 213–217. doi: 10.1177/0891988705281868

Manly, J.J., Touradji, P., Tang, M.X., & Stern, Y. (2003). Literacy and memory decline among ethnically diverse elders. *Journal of Clinical and Experimental Neuropsychology*, 25(5), 680–690. doi:DOI 10.1076/jcen.25.5.680.14579

Maruff, P., Lim, Y.Y., Darby, D., Ellis, K.A., Pietrzak, R.H., Snyder, P.J., ... Masters, C.L. (2013). Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC Psychology*, 1(1), 30. doi: 10.1186/2050-7283-1-30

Mielke, M.M., Machulda, M.M., Hagen, C.E., Edwards, K.K., Roberts, R.O., Pankratz, V.S., ... Petersen, R.C. (2015). Performance of the CogState computerized battery in the Mayo Clinic Study on Aging. *Alzheimers & Dementia*, 11(11), 1367–1376. doi: 10.1016/j.jalz.2015.01.008

Petersen, R.C., Lopez, O., Armstrong, M.J., Getchius, T.S.D., Ganguli, M., Gloss, D., ... Rae-Grant, A. (2018). Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*, 90(3), 126–135. doi: 10.1212/WNL.0000000000004826

Potter, G.G., Plassman, B.L., Burke, J.R., Kabeto, M.U., Langa, K.M., Llewellyn, D.J., ... Steffens, D.C. (2009). Cognitive performance and informant reports in the diagnosis of cognitive impairment and dementia in African Americans and whites. *Alzheimer's & Dementia*, 5(6), 445–453. doi:https://doi.org/10.1016/j.jalz.2009.04.1234

Rahman-Filipiak, A.M., Giordani, B., Heidebrink, J., Bhaumik, A., & Hampstead, B.M. (2018). Self- and Informant-Reported Memory Complaints: Frequency and Severity in Cognitively Intact Individuals and those with Mild Cognitive Impairment and Neurodegenerative Dementias. *Journal of Alzheimers & Disease*, 65(3), 1011–1027. doi: 10.3233/jad-180083

Scott, E.P., Sorrell, A., & Benitez, A. (2019). Psychometric Properties of the NIH Toolbox Cognition Battery in Healthy Older Adults: Reliability, Validity, and Agreement with Standard Neuropsychological Tests. *Journal of the International*

Neuropsychological Society, *25*(8), 857–867. doi: 10.1017/S1355617719000614

Shin, J. & Doraiswamy, P.M. (2016). Underrepresentation of African-Americans in Alzheimer's Trials: A Call for Affirmative Action. *Frontiers in Aging Neuroscience*, *8*, 123–123. doi: 10.3389/fnagi.2016.00123

Snitz, B.E., Bieliauskas, L.A., Crossland, A., Basso, M.R., & Roper, B. (2000). PPVT-R as an Estimate of Premorbid Intelligence in Older Adults. *The Clinical Neuropsychologist*, *14*(2), 181–186. doi: 10.1076/1385-4046(200005)14:2;1-Z;FT181

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, *8*(3), 448–460.

Stern, Y. (2006). Cognitive Reserve and Alzheimer Disease. *Alzheimer Disease & Associated Disorders*, *20*(2), 112–117.

Stricker, N.H., Lundt, E.S., Edwards, K.K., Machulda, M.M., Kremers, W.K., Roberts, R.O., . . . Mielke, M.M. (2019). Comparison of PC and iPad administrations of the Cogstate Brief Battery in the Mayo Clinic Study of Aging: Assessing cross-modality equivalence of computerized neuropsychological tests. *The Clinical Neuropsychologist*, *33*(6), 1102–1126. doi: 10.1080/13854046.2018.1519085

Sundermann, E.E., Maki, P.M., Rubin, L.H., Lipton, R.B., Landau, S., & Biegon, A. (2016). Alzheimer's Disease Neuroimaging Initiative. Female advantage in verbal memory: Evidence of sex-specific cognitive reserve. *Neurology*, *87*(18), 1916–1924. doi: 10.1212/WNL.0000000000003288

Sundquist, J., Winkleby, M.A., & Pudaric, S. (2001). Cardiovascular disease risk factors among older black, Mexican-American, and white women and men: an analysis of NHANES III, 1988–1994. Third National Health and Nutrition Examination Survey. *Journal of the American Geriatrics Society*, *49* (2), 109–116. doi: 10.1046/j.1532-5415.2001.49030.x. PMID: 11207863.

Weintraub, S., Besser, L., Dodge, H.H., Teylan, M., Ferris, S., Goldstein, F.C., . . . Morris, J.C. (2018). Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). *Alzheimer Disease and Associated Disorders*, *32*(1), 10–17. doi: 10.1097/wad.0000000000000223

Weintraub, S., Dikmen, S.S., Heaton, R.K., Tulsky, D.S., Zelazo, P.D., Bauer, P.J., . . . Gershon, R.C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*, S54–S64. doi: 10.1212/WNL.0b013e3182872ded

Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N.R., Chui, H., . . . Morris, J.C. (2009). The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. *Alzheimer Disease and Associated Disorders*, *23*(2), 91–101. doi: 10.1097/WAD.0b013e318191c7dd

Wild, K.V., Mattek, N.C., Maxwell, S.A., Dodge, H.H., Jimison, H.B., & Kaye, J.A. (2012). Computer-related self-efficacy and anxiety in older adults with and without mild cognitive impairment. *Alzheimers & Dementia*, *8*(6), 544–552. doi: 10.1016/j.jalz.2011.12.008

Wilkinson, G.S., Robertson, G.J., & Psychological Assessment Resources, I. (2006). *WRAT 4 : wide range achievement test; professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.O., . . . Petersen, R.C. (2004). Mild cognitive impairment - beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, *256*(3), 240–246. doi:DOI 10.1111/j.1365-2796.2004.01380.x

Zahodne, L.B., Sol, K., & Kraal, Z. (2019). Psychosocial Pathways to Racial/Ethnic Inequalities in Late-Life Memory Trajectories. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *74*(3), 409–418. doi: 10.1093/geronb/gbx113

Zuckerman, I.H., Ryder, P.T., Simoni-Wastila, L., Shaffer, T., Sato, M., Zhao, L., & Stuart, B. (2008). Racial and ethnic disparities in the treatment of dementia among Medicare beneficiaries. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *63*(5), S328–S333. doi: 10.1093/geronb/63.5.s328