6-1-2019

Dementia-Related Neuropsychological Testing Considerations in Non-Hispanic White and Latino/Hispanic Populations

Shanna L. Burke
Florida International University

Mitra Naseh
Portland State University, mitra.naseh@pdx.edu

Miriam Rodriguez
Carlos Albizu University

Aaron Burgess
Florida International University

David Loewenstein
University of Miami

Citation Details

Burke, S. L., Naseh, M., Rodriguez, M. J., Burgess, A., & Loewenstein, D. (2019). Dementia-related neuropsychological testing considerations in non-Hispanic White and Latino/Hispanic populations. Psychology & neuroscience, 12(2), 144.

HHS Public Access
Author manuscript
Psychol Neurosci. Author manuscript; available in PMC 2019 October 24.

Published in final edited form as:
Psychol Neurosci. 2019 June ; 12(2): 144–168. doi:10.1037/pne0000163.

Dementia-Related Neuropsychological Testing Considerations in Non-Hispanic White and Latino/Hispanic Populations

Shanna L. Burke,
Robert Stempel College of Public Health and Social Work, School of Social Work, Florida International University

Mitra Naseh,
Robert Stempel College of Public Health and Social Work, School of Social Work, Florida International University

Miriam J. Rodriguez,
PsyD Program, Carlos Albizu University

Aaron Burgess,
Robert Stempel College of Public Health and Social Work, School of Social Work, Florida International University

David Loewenstein
Center on Aging as the Center for Cognitive Neuroscience and Aging and Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami

Abstract

Hispanic individuals are at greater risk for health disparities, less than optimal health care, and are diagnosed at later stages of cognitive impairment than white non-Hispanics. Acculturation and different attitudes toward test-taking may result in decrements in performance, especially on unfamiliar measures that emphasize speed and accuracy. Non-Hispanic individuals often outperform Hispanic individuals on cognitive and neuropsychological measures in community and clinical populations. Current neuropsychological testing may not provide accurate data related to monolingual and bilingual individuals of Hispanic descent. Testing instruments were identified by searching academic databases using combinations of relevant search terms. Neuropsychological instruments were included if they were designed to detect cognitive impairment, had an administration time of less than 45 minutes, and were available in English. Validity studies were required to employ gold standard comparison diagnostic criteria. Twenty-nine instruments were evaluated in dementia staging, global cognition, memory, memory and visual abilities, working memory and attention, verbal learning and memory, recall, language, premorbid intelligence, literacy/cognitive reserve, visuospatial, attention, problem-solving, problem solving and perception, functional assessment, and mood/daily functioning domains. Spanish-language neuropsychological instruments need to be made widely available and existing instruments to be normed in Spanish to best serve and assess diverse populations. Psychometric data were reported

Correspondence concerning this article should be addressed to Shanna L. Burke, Robert Stempel College of Public Health and Social Work, School of Social Work, Florida International University, 11200 South West 8th Street, AHC5 585, Miami, FL 33199. sburke@fiu.edu.

for neuropsychological instruments, which may be administered to Hispanic older adults presenting for evaluation related to dementia-spectrum disorders. This is one of the few reviews to provide an overview of the sensitivity and specificity of available Spanish translated neuropsychological instruments.

Keywords

dementia; neuropsychological testing; sensitivity; Spanish; specificity

Hispanic individuals are at greater risk for health disparities and less than optimal health and mental health care (Langellier, Chen, Vargas-Bustamante, Inkelas, & Ortega, 2016; Vega, Rodriguez, & Gruskin, 2009). Individuals from these population groups require even greater diligence from health care providers to detect cognitive impairment or conditions such as delirium, which may be mistakenly diagnosed as mild cognitive impairment (MCI) or dementia (Siddiqi, House, & Holmes, 2006). The disparities known to exist in the recognition and diagnosis of dementia in Hispanics, support the concern that a delirium state will be under recognized and improperly diagnosed. Studies have indicated that similar to other ethnic minorities, Hispanics are diagnosed at later stages of cognitive impairment than white non-Hispanics (Chin, Negash, & Hamilton, 2011). Hispanics may experience higher risk factors that are associated with cognitive impairment and AD development, such as diabetes, cardiovascular risk factors (Evans et al., 1997; Haan et al., 2003; Lopez et al, 2003), and vascular disease, (O'Bryant et al., 2007, 2013; Tang et al., 2001), which may be mistaken for AD itself or cause delirium symptoms. Moreover, some evidence suggests that Hispanics present with higher levels of depressive symptoms and cognitive impairment than white non-Hispanics, which may contribute to the higher prevalence of AD diagnosis (Bell-McGinty et al., 2002; Livney et al., 2011), in people with and, most concerning, without AD pathology.

Individual cultural attitudes may affect neuropsychological, and to a lesser extent, functional test scores (Ostrosky-Solis, Ramirez, & Ardila, 2004; Rosselli et al., 2002; Sayegh & Knight, 2013; Tappen, Rosselli, & Engstrom, 2010). Cultural factors include differences how one might consider of the facilitators request to interpret and respond to a given task with the utmost speed, rigor, and precision (Ardila, 2013; Green, Rohling, Lees-Haley, & Allen, 2001). Acculturation level (Arnold, Montgomery, Castañeda, & Longoria, 1994; Razani, Burciaga, Madore, & Wong, 2007) and attitudes may influence test taking. Which may result in poorer performance It is established that effort can account for more variance than cognitive impairment (Green et al., 2001), and significant differences in testing outcomes can be accounted for by differences in levels of motivation (Liu, Bridgeman, & Adler, 2012). As such, effort and motivation remain crucial concepts to explore when considering differing approaches to test-taking. For instance, the requirement for providing optimal performance for a task may be deemphasized in certain cultures. This issue arises in all levels of testing, including high-stakes testing related to academic performance and college entrance exams. In one paper, Altshuler and Schmautz (2006) found that certain cultural values that are emphasized in Hispanic culture may, at times, appear to influence maximal performance. These values include interdependence over autonomy, and an allocentric self-concept as opposed to an idiocentric self-concept. As such, white non-Hispanic participants may have an advantage increasing performance on neuropsychological tests developed and delivered in English with instructions placing importance on autonomy and independence.

Current neuropsychological testing, utilizing traditional assessment instruments, may not be sufficient in providing accurate data related to disease diagnosis and progression in bilingual participants. In cross-cultural studies, the use of neuropsychological test scores, as an index of the severity of underlying disease process in the brain (even when using validly translated and back-translated tests), may be subject to nuances of language. One recent study found that while traditional neuropsychological instruments demonstrated linear patterns of decline in monolingual participants, a quadratic decline was found for bilingual individuals (Anderson, Saleemi, & Bialystok, 2017). This same study raised the question of whether classification errors may occur as a result of utilizing standard neuropsychological tests with bilingual participants. Previous studies have found that non-Hispanic individuals outperform Hispanic individuals in the dementia and clinical populations among cognitive screeners and other neuropsychological measures (Boone, Victor, Wen, Razani, & Pontón, 2007; Gross et al., 2015; Mungas, Reed, Farias, & DeCarli, 2009; Razani, Burciaga, et al., 2007; Razani, Murcia, Tabares, & Wong, 2007), even though, in some cases, volumes of brain regions among Hispanic individuals indicated less atrophy than White non-Hispanic participants in the same study Burke et al., 2018. The purpose of this study is to review prominent validated neuropsychological assessments within the published literature to determine efficacy for implementation with diverse English and Spanish speaking populations based on gold standard diagnostic criteria.

Methods

Neuropsychological instruments were identified by searching electronic databases (Entrez-PubMed, CINAHL, PsycINFO, EBSCOhost, and Google Scholar), using combinations of the following terms: “dementia,” “Alzheimer’s disease,” “cognitive impairment,” “screening,” “English,” “Spanish,” “sensitivity,” and “specificity.” Individual test names were also used as search terms. In addition, the reference lists of papers located in the original search were then manually explored. We did not evaluate global mental status tests, such as the mini-mental status exam (MMSE) and the Montreal Cognitive Assessment (MOCA) because there have been a number of reviews on cognitive screeners. Neuropsychological measures were included if they were employed in the assessment of cognitive impairment or had been used for that purpose, had an administration time of less than 45 minutes and were available in English. These instruments could be administered directly to patients, or be partially or fully informant informant-rated. Individual papers relating instrument information were included if they: were the original paper presenting the content of the test; presented data relating to the screening aspects of the instrument (as opposed to aspects such as factor structure, which are outside the scope of the current paper); presented data relating to the performance of the test as it stands alone (i.e., validity statistics based on scores from combined sources [screen test plus functional status, for example] were not considered).

Validity studies must also have employed acceptable “gold standard” diagnostic criteria (i.e., based on international diagnostic guidelines or clinical judgment following a full assessment battery); the use of another screening test as the gold standard was not acceptable. Where denoted, the test sensitivity is the ability of a test to correctly identify those with the disease (true positive rate). Test specificity is the ability of the test to correctly identify those without the disease (true negative rate). Parikh and colleagues (2008) provide an excellent overview of sensitivity, specificity, and positive and negative predictive values. It should be noted that sensitivity and specificity applies to the aspects of a test while actual positive and negative values are dependent on the percentage of true positive and true negative cases in a population.

Data were extracted for each screen by one author. A list of cognitive, psychiatric, and functional domains/abilities assessed by each test was created independently by two of the authors, and a final list was agreed upon by consensus, with a third author consulted as necessary. Neuropsychological tests were organized by classifications according to the domain structure set forth by McKhann et al. (2011). These domains included: dementia staging, global cognition, memory, memory and visual abilities, working memory and attention, verbal learning and memory, recall, language, premorbid intelligence, literacy/cognitive reserve, visuospatial, attention, problem-solving, problem solving and perception, functional assessment, and daily functioning. Table 1 displays the results of this study in the order in which they appear in the text.

Dementia Staging: Clinical Dementia Rating (CDR) Scale

The CDR (Morris, 1993) is a scale that assesses global staging in the development and severity of dementia (Juva et al., 1995). The CDR Sum of boxes (CDRsb) demonstrated 71% sensitivity and 81% specificity (cut-off score of 2.5 or higher) in distinguishing dementia, 74% sensitivity and 81% specificity (cut-off score of 2.5 or higher) in distinguishing probable Alzheimer’s disease (AD), and 80% sensitivity and 69% specificity (cut-off score of 2.0 or higher) in distinguishing possible AD in a sample extracted from the 2008 National Alzheimer’s Coordinating Center (NACC) uniform data set (O’Bryant et al., 2010). The CDR has been shown to be able to distinguish between participants with AD and controls when comparing English speakers and Spanish speakers. Even in the presence of education, age, and cultural differences, the CDR performed as intended in similar cohorts (Sano et al., 1997). Although a thorough review of the literature did not yield Spanish language CDR sensitivity or specificity, it should be noted that a linguistically adapted CDR instrument was developed in 2010 for Puerto Rican populations (Oquendo-Jiménez, Mena, Antoun, & Wojna, 2010).

Global Cognition: Alzheimer’s Disease Assessment Scale Cognitive Subscale (ADAS-Cog)

The ADAS-Cog is one of the most widely utilized cognitive assessment tools, measuring memory, language, praxis, attention, and cognition (Kolibas, Korinkova, Novotny, Vajdickova, & Hunakova, 2000). This tool was originally designed to detect mild cognitive impairment and mild to moderate AD (Podhorna, Krahnke, Shear, & Harrison, 2016). It has been reported that the ADAS-Cog is limited in its ability to accurately measure progression of cognitive impairment when the original scoring protocols are used in comparison with a newer scoring method based on a revised factor analysis (Benge, Balsis, Geraci, Massman, & Doody, 2009). In addition, researchers found that four items have measurement bias, which has resulted in substantial differences in the answers between men and women (Verma et al., 2015). Another study found that the items on the ADAs-Cog were usually too easy for participants, and this created a very large ceiling affect (50%; Cano et al., 2010). This test is widely used in patient settings and clinical trials (Hobart et al., 2013), but has many variations in scoring and protocols for administration (Connor & Sabbagh, 2008). This prompted the revision and expansion of the ADAS-Cog to the development of the Alzheimer's disease Assessment Scale-Cognitive-Plus (ADAS-CogPlus; Skinner et al., 2012), which was created to improve responsiveness for MCI measurement and added functional ability and executive function domains. However, a thorough review of the literature did not yield ADAS-Cog-Plus sensitivity or specificity data. The ADAS-cog (cut-off score of 10) yielded 78% sensitivity and 100% specificity in a sample of 75 older adults (45 MCI vs. 30 controls; Perneczky et al., 2006). The Spanish version of this assessment scale (cut-off score of 10) yielded 95.5% sensitivity and 72.94% specificity in distinguishing AD in a study of 451 (254 controls, 86 with MCI, and 111 with AD) individuals in Spain (Monllau et al., 2007). The adjusted Spanish version of the ADAS-cog for age and schooling with the cut-off score of 12 yielded 89.19% sensitivity and 88.53% specificity (Monllau et al., 2007). Some challenges for Hispanic populations on this measure may include the culturally biased items. For example, items used in the word recall and naming subtests may be less salient for some Hispanic cultures due to less exposure to those specific words and items.

Memory

The Fuld Object Memory Test (FOME or Fuld).—The FOME or Fuld Test (Loewenstein, Duara, Argüelles, & Argüelles, 1995) has been validated as a culture-fair method to assess impairment of episodic memory. Objects included in this test were selected on the basis of minimizing cultural bias. This measure has specifically shown high cultural validity in MCI and later stages of AD (Loewenstein et al., 1995). In a multilanguage study, the FOME (cut-off score of 29 or less) yielded 95.9% sensitivity and 100% specificity for Spanish speakers (AD = 27 vs. controls = 23) and 95.5% sensitivity and 96.7% specificity in English-speaking patients (AD = 111 vs. controls = 30; Loewenstein et al., 1995). The FOME (cut-off score of 30) yielded 93.2% sensitivity and 63.5% specificity in a study among 140 outpatients (88 with dementia and 52 controls) from the Detroit satellite of the Michigan Alzheimer's disease Research Center (MADRC; Mast, Fitzgerald, Steinberg, MacNeill, & Lichtenberg, 2001).

The Short-Term Visual Memory Binding Test.—The Visual Memory Binding Test (previously referred to as the Memory Capacity Test) examines short-term memory (STM) by asking participants to remember objects, colors, or the colors of objects. After trying to integrate these items into their memory, the participants are asked to recall the items, colors, or colors of items verbally. Participants with AD perform worse when asked to recall objects, and the objects and their respective colors, though their performance was significantly impaired when asked to remember objects and colors together (Parra et al., 2009). Recently this test has been shown to differentiate individuals with amnestic mild cognitive impairment and dementia from those who are cognitively normal (Mowrey et al., 2017). A recent investigation by Buschke et al. (2017) used a Total Number of Items recalled in the Paired condition (TIP) score of ≤22 for distinguishing amnestic MCI from controls (sensitivity = 0.74, specificity = 0.73) and amnestic MCI and dementia combined from controls (sensitivity = 0.84, specificity = 0.73) among 297 older adults (20 with dementia, 31 with amnestic MCI, and 246 controls). A 2017 study, which was a substudy of the community-based Einstein Aging Study, used a TIP score of ≤17 to distinguish dementia from amnestic MCI and controls (sensitivity = 0.95, specificity = 0.87). Although important for other tests, age and education adjustments did not have a significant effect in improving validity (Buschke et al., 2017). The Memory Binding Test (MBT) has been developed in Spanish and Catalan, and those two versions are considered equivalent (Gramunt et al., 2016), though the Spanish version was found to be affected by factors such as age, education, and sex (Gramunt et al., 2015). An Argentine version of the MBT has also been developed with words specific to that semantic context, and has shown high sensitivity (69%) and high specificity (88%) among a sample of 88 (46 controls and 42 with amnestic MCI) monolingual Rioplatense-Spanish speakers for detecting MCI (Roman et al., 2016).

Loewenstein-Acevedo Scales for Semantic Interference and Learning (LASSI-L).—The LASSI-L (Crocco, Curiel, Acevedo, Czaja, & Loewenstein, 2014) is a recently developed scale for the measure of cued recall that allows for the discernment of proactive and retroactive interference effects when global memory impairment is a control (Crocco et al., 2014). This test has yielded 87.9% sensitivity (when results of the first and second semantically related lists of word entered to the model) and 89.4% (when results of the first semantically related lists of word entered to the model) to 91.5% (when results of the second semantically related lists of word entered to the model) specificity in distinguishing between amnestic MCI and controls among a sample of 121 (47 normal, 34 with amnestic MCI, and 40 with probable AD) older adults. Primary language of the 70.8% of the normal subjects was English, and this percentage was 67.6% among older adults with amnestic MCI (Crocco et al., 2014). A recent study by Loewenstein and colleagues (Loewenstein et al., 2016) measured degrees of cognitive impairment in 93 participants utilizing the LASSI-L (the LASSI with an included learning assessment). The results of this study yielded deficits of 89% for those with MCI and 13% for individuals without impaired cognition. Additionally, this test was able to discern subtle differences in cognition to recognize deficits of 47% for this who were pre-MCI and 33% for individuals with subjective memory impairment (Loewenstein et al., 2016).

This sensitivity in recognizing subtle cognitive impairments through the LASSI-L, along with identifying increased amyloid load among neuropsychologically normal community-dwelling older adults, is a distinguishing characteristic of this assessment in relation to other comparable available tests (Loewenstein et al., 2015, 2016). Failure to recover from semantic interference (frPSI) has also been found to be related to related to volumetric loss or loss of cortical thickness in persons with MCI (Loewenstein et al., 2016, 2017).

More recently, the LASSI-L has been validated in Spanish-speaking populations, yielding an area under the curve for discriminating between healthy controls and aMCI equal to 0.909, and between healthy controls and mild AD equal to 0.98 (Matías-Guiu et al., 2017). This assessment tool yielded a sensitivity of 81.8%, 90.1%, and 75.7% (for cued recalls using list A and B of words, and delayed recall respectively) and specificity of 81.6%, 64.7%, and 92.9% (for cued recalls using list A and B of words, and delayed recall respectively) in distinguishing between participants with aMCI and healthy controls (Matías-Guiu et al., 2017). Utilizing frPSI semantic intrusions, the test is able to distinguish between middle-aged Spanish-speaking offspring of late onset AD patients and controls as well as different patterns of functional connectivity on fMRI. The Spanish versions include items that have minimal cultural bias, making it a stronger measure of semantic interference among Spanish-speaking populations.

Memory and Visuospatial Abilities: Brief Visuospatial Memory Test—Revised (BVMT-R)

The BVMT-R is an assessment tool that measures delayed recall and delayed yes/no recognition task (Benedict, Schretlen, Groninger, Dobraski, & Shpritz, 1996). The final trial of the BVMT-R may be particularly challenging for test-takers in cultures in which exposure to recognition tasks was minimal in the country where their primary and secondary education was obtained. This test is delivered in three trials, in which the respondent studies the stimulus page for 10 seconds and is asked to draw as many of the figures viewed as possible in their correct location on a page in the response booklet. After a 25-min timed hiatus, a delayed recall trial begins, in which the participant is asked to repeat the first task. In the third and final trial, the participant is asked to view figures and identify which were among the 12 figures that were included in the original trial. If desired, a fourth trial can be administered which involves copying the geometric figures. This can be used to screen for severe visuo-constructive deficits. The BVMT-R has yielded a sensitivity of 98% and a specificity of 82% in distinguishing older adults with dementia ($n = 45$) from controls ($n = 59$; Benedict et al., 1996). Moreover, this test (cut-off score of 17) yielded a sensitivity of 95% and specificity of 93% in distinguishing normal cognitive performance among a sample of 515 adults in the United States (Beier et al., 2017). A thorough review of the literature did not reveal Spanish language sensitivity and specificity data. However, Cherner and colleagues (2007) utilized this test with monolingual Spanish-speaking Mexican individuals and yielded preliminary data regarding further adaption recommendations. This included an accounting for years of education that yielded significant improvement in test findings as well considering the impact of acculturation on future test outcomes (Cherner et al., 2007).

Working Memory and Attention

Wechsler Memory Scale Fourth Edition (WMS-IV).—The WMS-IV (Pearson, 2009) is a widely used working memory scale. The current revised WMS consists of seven subtests to measure across memory indexes (Pearson, 2009). The Symbol Span subtest of the WMS-IV (cut-off score of 14) yielded sensitivity of 50% and specificity of 83% detection of Poor Effort among a sample of 143 patients (Young, Caron, Baughman, & Sawyer, 2012). Additionally, immediate memory versus delayed memory indexes were determined to yield 96% sensitivity and 87% specificity in determining multiple sclerosis patients versus controls in a sample of 40 (20 with relapsing-remitting multiple sclerosis patients and 20 controls) patients (Spedo, 2014). The WMS-IV and the third edition (WMS-III) of this tool are available in Spanish and commonly utilized in clinical settings. The WMS-III had a sensitivity of 71% and 89% in identifying Immediate and Delayed Memory Dysfunction respectively among patients with AD (Wechsler, 2004). In a study by Demsky and colleagues (1998) among 50 normal Hispanic Americans, using translated version of the WMS Revised resulted in a score with an average of one standard deviation lower than average. Therefore, it was suggested that clinicians should not use the translated version of the English language test without norming and validity testing (Demsky et al., 1998).

Wechsler Adult Intelligence Scale (WAISIV)-Digit Span and Letter Number Sequencing Tests.—The subtests Digit Span forward and backward, as well as letter number sequencing from the Wechsler Adult Intelligence Scale, fourth edition (WAIS-IV), are often used as screening measures for working memory performance. The Digit Span subtest of the WAIS, Revised was validated and with a cutting score of eight yielded a specificity of 100% and a sensitivity of 27% (Inman & Berry, 2002). Moreover, the Reliable Digit Span derived from the Digit Span subtest of the WAIS, with a cutting score of seven, yielded a specificity of 82% and a sensitivity of 39% among a sample of 138 veterans recruited in a traumatic brain injury clinic (Spencer et al., 2013). There have been Spanish language Digit Span adaptations for Mexican, Central American, and Puerto Rican populations, which reportedly yielded no major differences in Forward or Backward test results (Mejia, Hernandez, Lindsey, Daughtry, & Puente, 2014). Of note, the Spanish versions of Wechsler tests are reportedly difficult to obtain in the United States and there may be limits on the ability to market and purchase in the U.S. (Ferraro, 2015).

Verbal Learning and Memory

Verbal learning and memory tests are commonly used in clinical setting as a screening tool for MCI and dementia. These measures generally contain word lists that belong to certain categories (i.e., fruits, vegetables, or means of transportation). The list is presented over several trials to give the examinee an opportunity for learning. There is generally an immediate and delayed recall trial, as well as a recognition trial. Possible cultural factors that may interfere with optimal performance on these tests include the presence of culturally biased words, and an unfamiliarity among Hispanic patients with recognition trials.

Hopkins Verbal Learning Test – Revised (HVLT-R)

The HVLT-R is designed to test (Hogervorst et al., 2002) memory and is advantageous to the MMSE in that it has no ceiling effect nor education bias. This test has demonstrated 87% sensitivity and 98% specificity when assessing for dementia (cut-off score of 14.5) and 91% sensitivity and 98% specificity when assessing for AD (cut-off score of 24.5) in a sample of 82 demented patients and 114 controls as part of the Oxford Project To Investigate Memory and Ageing (OPTIMA; Hogervorst et al., 2002). A Spanish language HVLT-R assessment (in delay recall score <4) yielded 88% sensitivity and 70% specificity for amnestic MCI and 96% sensitivity and 85% specificity for AD (cutting point <13) in a sample of 298 older adults (54 with AD, 132 amnestic MCI, and 109 controls) recruited in community centers and a memory clinic in Spain (Gonzalez-Palau et al., 2013).

Recall: The Rey Auditory Verbal Learning Test (RAVLT)

The RAVLT (Schmidt, 1996) is a widely used assessment for short-term memory and learning measurement (Britt, Adams, Godding, Grothues, & Varnado, 1995). The RAVLT has demonstrated 75.7% sensitivity and 91.5% specificity in measuring true recognition, implicit memory, and automatic memory whereas true and primary memory recognition were associated with 73.8% sensitivity and 90% specificity in identifying noncredible memory performance in a sample of 174 (61 noncredible, 88 clinic patients, 25 controls) fluent English speakers (Boone, Lu, & Wen, 2005). A Spanish language study yielded a total score sensitivity of 65% and a specificity of 76% for immediate and delayed recall in a sample of 106 Hispanic adults in Puerto Rico (Neblina, 2012).

Language

Boston Naming Test (BNT).—The BNT (Kaplan, Goodglass, Weintraub, & Goodglass, 1983) is an assessment that tests for recalled naming ability in response to visual stimuli. Some cultural factors that may interfere in this test are the presentation of items that are not easily translated to Spanish or are culturally biased. The test has been found to be sensitive to performance changes along repetitive testing as well as unbiased by practice effects (Huff, Collins, Corkin, & Rosen, 1986). In a 2015 study, the BNT demonstrated 57% sensitivity and 59% specificity in predicting left seizure focus among a sample of patients with temporal-lobe epilepsy (Umfleet et al., 2015). However, two Spanish studies have identified 39% sensitivity and 89% specificity for detecting AD in a sample of 59 (23 with AD and 36 controls; Fernández & Fulbright, 2015) and 85% sensitivity and 94% specificity when assessing Spanish speakers with AD using the shortened BNT in a sample of 246 participants (103 with probable AD and 143 controls; Serrano et al., 2001).

Craft Story 21 Recall Immediate/Delayed.—The Craft Story 21 Recall Immediate is a test which consists of reading a short story to the participant, who is then asked to recall it from memory immediately and tell the interviewer. The primary measure of performance is the number of story units recalled and is scored in either a verbatim format or a paraphrased format. The delayed format requires the participant to tell the story to the interviewer later on in the interview and tests episodic memory. These assessments were utilized in a landmark study that demonstrated the impact of plasma insulin levels on patients' memory. This study determined that raising the level of insulin within the body helped to keep levels of plasma glucose lowered, therefore promoting memory enhancement within patients (Craft et al., 1996). The Craft Story 21 recall, both immediate and delayed, was utilized to measure the efficacy of this insulin treatment. Although sensitivity and specificity were not directly reported within this study, this assessment has significance in examining how insulin can impact memory functioning. Craft Story 21 was recommended to replace Logical Memory, Immediate and Delayed in the data collection battery for the National Alzheimer's Coordinating Center (NACC). Spearman's correlation coefficients were calculated between the pairs showing $r = .073$ between logical memory IA (immediate) and Craft Story 21 immediate (paraphrase), and $r = .77$ between logical memory IIA and Craft Story 21 delayed (paraphrase). The NACC study showed that both logical memory and Craft Story 21 had the lowest accuracy in prediction out of the tests examined in the neuropsychological battery (Monsell et al., 2016).

Controlled Oral Word Association Test (COWAT).—The COWAT (Borkowski, Benton, & Spreen, 1967) measures letter and category fluency in patients with MCI but is not sensitive enough to distinguish healthy controls from those with single-domain amnestic MCI (Malek-Ahmadi, Small, & Raj, 2011). This test (cut-off score 5) has yielded a sensitivity of 67% and specificity of 88% for detecting malingered neurocognitive dysfunction in a sample of 101 (46 presenting malingered neurocognitive dysfunction [MND] and 55 without MND) participants (Johnson, Silverberg, Millis, & Hanks, 2012). Moreover, this test showed sensitivity of 58% and specificity of 92% in a sample of 26 adults with Attention Deficit Hyperactivity Disorder (ADHD) and 26 controls (Lovejoy, 1998), and with the cut of score of 19, sensitivity of 36.5% and specificity of 89.3% among sample of. 969 veterans (Sugarman & Axelrod, 2015). A thorough review of the literature did not yield any Spanish language data regarding sensitivity or specificity of the COWAT. Cherner et al. (2008) administered the COWAT in Spanish (P-M-R) and English (F-A-S) and found no significant differences between the two groups.

Multilingual Naming Test (MiNT).—The MiNT is an assessment tool that is designed for use with bilingual patients. The purpose of the MiNT is to assess picture naming skills for the detection of naming impairments in patients with MCI and AD (Ivanova, Salmon, & Gollan, 2013). Items and administration procedures were designed to minimize cultural bias. Although this test has been carefully designed for assessment in different languages (i.e., English, Spanish, Chinese, and Hebrew), normative data and data on the validity of this test are scarce (Fernández, 2013). This test demonstrated positive results for bilingual assessment in patients with MCI and AD, with a sensitivity range of 80% to 88% in bilinguals, in a sample of 130 (68 with probable AD, 18 with amnestic MCI, and 44 controls) monolingual English speakers and 29 (18 with probable AD and 11 controls) Spanish–English bilinguals (Ivanova et al., 2013). The Multilingual Naming Test was recommended to replace the Boston Naming Test in the data collection battery for the National Alzheimer's Coordinating Center. Spearman's correlation coefficients were calculated between the pairs showing $r = .076$ (Monsell et al., 2016).

Premorbid Intelligence: Nelson Adult Reading Test—American English Version (AMNART)

The AMNART (Grober & Sliwinsk, 1991) is an English-language standardized education-based assessment that estimates verbal intellectual ability, premorbid intelligence and overall cognitive functioning (Lowe & Rogers, 2011). The AMNART has demonstrated 83% sensitivity and 81% specificity in dementia differentiation (Vanderploeg, 2014).

Literacy/Cognitive Reserve

Wide-Range Achievement Test-3rd (WRAT-3).—The Wide-Range Achievement Test is a test for English speakers. It contains three subtests: reading, spelling, and mathematics computation. The reading subtest is used to determine both literacy and cognitive reserve. A thorough review of the literature yielded a 46.2% (reading grade-level) and 55.8% (self-reported education) sensitivity and 84.1% (reading grade-level) and 77.3% (self-reported education) specificity for WRAT-3 in combined non-Hispanic White ($n = 51$) and non-Hispanic African American ($n = 62$) English-speaking adults (Rohit et al., 2007). Among non-Hispanic White participants, correction based on years of education yielded a higher sensitivity and specificity (Rohit et al., 2007). Among African Americans, correction based on reading scores yielded in greater specificity compared to correction derived from years of education (84.1% and 77.3% respectively; Rohit et al., 2007). Further, the WRAT-3 showed a diagnostic accuracy when measuring neurological impairment of African Americans with 48.4% sensitivity and 77.8% specificity, and 42.9% (47.6: when grade attainment was used as a correction factor) sensitivity and 88.5% (92.3: when grade attainment was used as a correction factor) specificity for White non-Hispanic participants (Rohit et al., 2007).

The fourth edition of the Wide-Range Achievement Test was published in 2006 (WRAT-4). Changes to this version include the addition of a sentence comprehension subtest, as an entirely new measure of reading achievement. Additionally, the age-based norms are extended from age 75 to 94. The reading subtest has been extended from 42 to 55 items and contains new words selected from the EDL Core Vocabularies in reading, mathematics, science, and social science. This was done to ensure that a sampling of new words was obtained from various grade levels (Wilkinson & Robertson, 2014). A thorough literature search revealed no studies examining sensitivity and specificity for dementia groups using this version of the test, nor Spanish language versions.

Word Accentuation Test (WAT).—The WAT measures premorbid intelligence of Spanish speaking patients through an assessment of the accentuation of infrequently used Spanish words that are written without accentuation denotation (Del Ser et al., 1997). This test has yielded a 78% sensitivity and 82% specificity in assessing for MCI (Del Ser et al., 1997).

Visuospatial

Rey–Osterrieth Complex Figure (ROCF)

The ROCF is a widely used assessment tool that evaluates visuospatial construction ability, visual memory, visual-motor integration skills (Davies, Field, Andersen, & Pestell, 2011), and nonverbal memory (Frank & Landeira-Fernandez, 2008). This test yields 80% sensitivity and 90% specificity in measuring neurocognitive response bias (cut-off of ≤50) in a sample of 146 credible and 157 noncredible patients (Reedy et al., 2013). Additionally, in a study where participants were older adult Spanish-speakers from Spain, the ROCF demonstrated 85.7% sensitivity and 98.3% specificity in long delay cued recall and 70% sensitivity and 62% specificity for the execution time of Rey Figure in predicting conversion to dementia after an MCI diagnosis (García-Herranz, Díaz-Mardomingo, & Peraita, 2016). This last factor has a timing component which may interfere with performance among cultural groups who have not been previously exposed to this type of testing.

Attention

Stroop Color and Word Test (SCWT).—The SCWT (Golden, 1978) is a measure of executive functioning. It is shown to be sensitive to measuring cognitive dysfunction and is a tool to measure a patient's ability to identify the color of a printed word while ignoring the content of the written word itself (Mackin, Ayalon, Feliciano, & Areán, 2010). An age and education corrected scaled score of this test has yielded an 88% sensitivity and 36.8% specificity (cutoff score of 7) in a sample of 52 older adults recruited at a community mental health agency in San Francisco (Mackin et al., 2010). Moreover, this test showed 59.2%

sensitivity and 57.8% specificity with the cutoff score of 5 (Mackin et al., 2010). Evidence was located to determine that in a monolingual/bilingual English and Spanish (40 English monolinguals, 11 Spanish monolinguals, and 71 Spanish–English bilinguals) study in South Florida, bilinguals had a slower response time to the SCWT than their monolingual counterparts (Rosselli et al., 2002). However, a 2014 study of adult native Spanish bilingual individuals from the United States–Mexico border region yielded findings of a faster response time to the Golden version of the Stroop test and subsequent inhibitory control advantage for those respondents with greater bilingual proficiency (Suarez et al., 2014). Using this test among 200 Spanish-speakers (100 diagnosed with ADHD and 100 controls) to distinguish ADHD demonstrated 81% sensitivity and 72% specificity (LópezVillalobos et al., 2010).

Trail Making Test.—The Trail Making Test A and B (TMT-A & TMT-B; Reitan, 1958) are measures of a patient's sequencing, psychomotor, visuomotor speed, spatial tracking, cognitive flexibility, and set shifting ability. Note that set shifting ability is a part of executive functioning that allows a person to switch from one area of concentration to another (Ravizza & Carter, 2008). Cultural factors that may interfere with performance on this measure among Hispanic individuals is the timing component, which may be affected by cultural differences in test-taking attitudes toward speedy completion of the measure. The TMT-A is reported to yield a 72% sensitivity and 82% specificity for suboptimal effort in a sample of 76 (58 with optimal effort and 18 with suboptimal effort) brain injury patients (Powell, Locke, Smigielski, & McCrea, 2011). Further studies have indicated 48% sensitivity and 85% specificity for TMT-A & B in a sample of 413 adults (Busse & Whiteside, 2012). Additionally, a study measuring traumatic brain injury (causing cognitive impairment) indicated between 87% and 100% specificity for TMT-A in a sample of (n = 42 suspected malingerers and n = 77 genuine) patients with very mild head injury (Iverson, Lange, Green, & Franzen, 2002). The TMT-A yielded sensitivity of 24.0% and specificity of 88.0% among monolingual Spanish-speaking Hispanic participants of a study among people living with HIV in distinguishing HIV-Associated Neurocognitive Disorders (Seay, 2015). Performance on the TMT yields two scores: times to completion (in seconds) for Parts A and B. Additionally, derived scores (i.e., difference B – A, and ratio B:A) are oftentimes used in clinical practice to remove the speed component from the test performance, provide a more pure measure of executive control, and serve as a possible symptom validity indicator (Arango-Lasprilla et al., 2015).

Problem-Solving: Trail Making Test B (TMT-B)

The TMT-B yielded a 50% sensitivity and 80.3% specificity for suboptimal effort in a sample of 76 brain injury patients (Powell et al., 2011). This test has demonstrated 69.0% sensitivity and 66.9% specificity in detecting mild cognitive impairment, 80.7% sensitivity and 31% specificity in detecting AD, and 71.6% sensitivity and 66.9% specificity in measuring for both mild cognitive impairment and AD (Ashendorf et al., 2008). A survey of 475 neuropsychologists in the United States found that the TMT-B is one of the 10 most often utilized tests with bilingual and monolingual Hispanic participants in the United States (Echemendia & Harris, 2004). A 2011 study determined that, for native Spanish speakers, versions of the TMT-B that utilized "Ch" and "D" sounds produced comparable and equivalent results (Cherner et al., 2008). The TMT-B yielded sensitivity of 60.0% and specificity of 64.0% among monolingual Spanish-speaking Hispanic participants of a study among people living with HIV (Seay, 2015).

Problem-Solving and Perception: Clock Face Drawing

The Clock Face Drawing Test (Tuokko, Hadjistavropoulos, Miller, & Beattie, 1992) is an assessment that seeks to measure a patient's frontal and temporal-parietal functioning. The Clock Face Drawing Test has been studied extensively, though the Clock Test, which uses a predrawn clock, is commonly used in older adult populations. A clock drawing type test is included in the Mini-Cog, MMSE, and MOCA assessments. This assessment has yielded a sensitivity of 86% and specificity of 92% in classifying older adults with AD versus controls (Agrell & Dehlin, 1998). Similar sensitivity rates with specificity of 96% were also reported for this test (Burns, Lawlor, & Craig, 2002). The ability to accurately detect an issue in cognition varied by level of impairment but not by patient status; classifying 78% of older adults with normal cognition, 89% with a mixed dementia type, and 77% of individuals with Ad (Manos & Wu, 1994). Among 31 patients with AD and 31 controls in Peru the Spanish version of the test using a cut-off score of 6 demonstrated 83.9% sensitivity and 93.5% specificity (Oscanoa, 2004).

Functional Assessment

Functional Activities Questionnaire (FAQ).—The FAQ (Pfeffer, Kurosaki, Harrah, Chance, & Filos, 1982) measures social functioning and activities of daily living, and mild cognitive impairment (Teng et al., 2010). This tool is useful to monitor the changes of functioning in activities of daily living that require higher cognitive ability (Mayo, 2016). The FAQ yielded 80% sensitivity and 87% specificity for distinguishing mild cognitive impairment and very mild AD progression in a sample of 1,801 individuals from the NACC dataset (1,108 with MCI and 693 with AD; Teng et al., 2010). Further, there were reported 85% sensitivity in identifying functional impairment (Mayo, 2016). Cultural factors that may affect scores on this measure include differences in social/cultural habits, family structure, and lifestyle. However, a 2010 study of 691 individuals indicated no differences in sensitivity and specificity outcomes between non-Hispanic and Hispanic participants (Tappen et al., 2010). A Spanish language validation yielded 71% (for illiterate participants) and 36% (for participants with 1–4 and 5–9 years of education) sensitivity and 79% (for illiterate participants), 89% (for participants with 1–4 years of education), and 98% (for participants with 5–9 years of education) specificity across education levels in 3,934 older adults (55 with mild to moderate dementia, 74 mild MCI, 185 controls) in Mexico City (Mejia, Gutiérrez, Villa, & Ostrosky-Solís, 2004).

Blessed Dementia Rating Scale (BDRS).—The BDRS provides informant-based data from a simple daily functional scale. It is advantageous in that it allows for long-term evaluation, of which the MMSE is unable to perform (Brickman et al., 2002). The BDRS comes in both full and modified form, where the full test assesses for everyday activities, habits, personality, interest, and drive, whereas the modified version simply tests for everyday activities and habits (Park, Jung, & Lim, 1995). The BDRS yielded 96% sensitivity and 82% specificity in distinguishing subjects with dementia among a sample of 291 older adults (Heun, Papassotiropoulos, & Jennssen, 1998). These activities and habits may significantly differ between Hispanic and non-Hispanic cultures. The BDRSTotal (cut-off score of 3.5) demonstrated 87% sensitivity and 90% specificity whereas the BDRS-Mod (cut-off score of 1.5 for sum of changes in everyday activities and changes in habits) demonstrated 90% sensitivity and 89% specificity for detection of AD in a sample of 451 (86 with MCI, 111 with AD, 254 controls) Spanish-speakers (Peña-Casanova et al., 2005). A thorough review of the literature did not yield any additional language data regarding sensitivity and specificity. A comparable scale to the BDRS but with information on a Spanish version is the Bayer Activities of Daily Living scale (B-ADL), which is a functional scale used for those with mild cognitive impairment and mild AD. The B-ADL demonstrated 81% sensitivity and 72% specificity to detect differences between MCI and AD, and correlates well with the Blessed Dementia Rating Scale ($r = .7$; Sánchez-Benavides et al., 2009).

The Direct Assessment of Functional Status Scale (DAFS).—The DAFS (Loewenstein et al., 1989) was one of the first performance-based scales for AD and has been translated into multiple languages. The DAFS utilizes caregiver reports to determine patient dementia progression across five daily living task domains—driving, telling time, remembering lists, basic financial tasks, and complex financial tasks (Pereira, Oliveira, Diniz, Forlenza, & Yassuda, 2010). A study on the criterion validity of the financial skills subscale of the DAFS yielded 75% sensitivity and 96% specificity in classifying those who manage their finances independently from those who do not (Barrett et al., 2009). A thorough review of the literature did not yield sensitivity and specificity data for Spanish studies. However, a landmark Brazilian Portuguese adaption of this scale yielded 80.6% sensitivity and 84.4% specificity for discerning MCI and 100% sensitivity and 93.7% specificity for discerning AD in the patient sample (N 89; Pereira et al., 2010). The DAFS-Revised was developed in response to the need to more sensitively measure early signals and changes in MCI in what the DAFS was currently not capturing (McDougall, Becker, Vaughan, Acee, & Delville, 2010). The DAFS-R demonstrated cross-cultural validity with a Brazilian sample of 89 old adults, yielding 100% sensitivity and 93.7% specificity in identifying AD and 80.6% sensitivity and 84.4% specificity for mild cognitive impairment (Pereira et al., 2010).

Mood/Daily Functioning

Geriatric Depression Scale (GDS).—The GDS (Yesavage et al., 1982) assesses for behavioral clinical depression in older adults through self-report (Marc, Raue, & Bruce, 2008). A review of the literature yielded a sensitivity of 100% and a specificity of 94% in the detection of major depression (cut-score of 14) and a sensitivity of 100% and a specificity of 91% in major/double depression cut-score of 13) in a sample of 119 patients (Low & Hubley, 2007). The experience of depressive symptoms has been noted to differ across cultures. For example, Hispanic cultures often times present with more somatic symptoms. A study on 192 older adult Spanish speakers in Spain demonstrated 86.7% sensitivity and 63.1% specificity to detect depression (Fernández-San Martín et al., 2002).

Center for Epidemiologic Studies Depression Scale (CES-D) Revised.—The CES-D Revised (Eaton, Smith, Ybarra, Muntaner, & Tien, 2004) is a widely used self-report depression inventory that measures depressive symptomology (Carleton et al., 2013). This scale (cut-off point of 16) has demonstrated a sensitivity of 86% and specificity of 90% in distinguishing poststroke depression among a sample of 80 stroke patients (Parikh, Eden, Price, & Robinson, 1989). Additionally, a study utilizing the Spanish version of the CES-D among a sample of 194 participants (70 with major depressive episode, 63 without major depression but with clinical diagnosis of other psychiatric disorders, and 61 with no evidence of psychiatric disorders) has yielded a sensitivity of 77.1%, specificity of 79.4%, and 78.2% of participants correctly classified with major depressive disorder or other psychiatric disorders with the cut-off score of 29 (Ruiz-Grosso et al., 2012). For those with major depressive disorder or no evidence of any psychiatric disorders, the Spanish scale (cut-off score of 24) demonstrated a sensitivity of 91.4%, a specificity of 96.7%, and an ability to correctly classify in 93.9% of cases with a major depressive episode (Ruiz-Grosso et al., 2012). Further evidence of the efficacy of the CES-D with diverse populations is demonstrated in the following. This test demonstrated very high reliability, $\alpha = .94$, in measuring depression with Hispanic women who were exposed to HIV and intimate partner violence risk (Mitrani, McCabe, Gonzalez-Guarda, Florom-Smith, & Peragallo, 2013). A 2011 study of 504 Spanish-speaking women demonstrated CES-D reliability of $\alpha = .93$ for women with low levels of acculturation and .94 for women with high levels of acculturation (McCabe, Vermeesch, Hall, Peragallo, & Mitrani, 2011).

Neuropsychiatric Inventory Questionnaire (NPI-Q).—The NPI-Q (Cummings et al., 1994) is one of the most widely used tools for outcome measures regarding the evaluation of behavioral disturbances in patients with dementia (Lai, 2014). This tool has demonstrated a sensitivity of 74.1% and a specificity of 79.5% in assessing neuropsychiatric symptoms in a sample of 173 patients with stroke or transient ischemic attack (Wong et al., 2014). Further, a validation of the Spanish translation of the NPI-Q, known as the NPI (Boada, Cejudo, Tàrraga, López, & Kaufer, 2002), has demonstrated success in screening both patient symptomology ($r = .879$) and as well as caregiver distress ($r = .92$; Boada et al., 2002).

Discussion

In total, 29 instruments were included in this study for evaluation (see Table 1 for results). The following post facto domains emerged: dementia staging, global cognition, memory, memory and visuospatial abilities, working memory and attention, verbal learning and memory, recall, language, premorbid intelligence, literacy/cognitive reserve, visuospatial orientation, attention, problem-solving, problem-solving and perception, functional assessment, and mood/daily functioning. Important findings emerged within these domains. Around half of the domains had assessment tools with data on gold standard outcomes for both English- and Spanish-speakers. Among the assessment tools with data on both Spanish- and English-speakers, only four assessments were found with consistent outcomes above 80% (FOME or Fuld, HVLT-R, Clock Face Drawing, and BDRS). These assessments were in the following domains: memory, verbal learning and memory, problem-solving and perception, and functional assessment (see Table 1). The assessment tools for memory, visuospatial abilities, and premorbid intelligence also had consistent outcomes exceeding 80%, but lack data on sensitivity and specificity among Spanish-speakers. The assessment tools in dementia staging and global cognition demonstrated global outcomes above 70%, but the CDR lacked data on Spanish-speakers. The assessments that incorporated memory tests demonstrated a wide range of cross-language outcomes. The FOME or Fuld demonstrated the highest outcomes across language with all scores in a range of 95–100%, whereas scores as low as 27% were reported for sensitivity of WAIS-IV among English-speakers. Cross-language outcomes were limited on assessments related to language domain, except for the BNT, which demonstrated low sensitivity and specificity. The assessment tool under the premorbid intelligence domain was only applicable for English-speaking populations and demonstrated outcomes above 80%. Cross-language data outcomes were not applicable to tools related to literacy domains, and outcomes had a wide range. Similarly, outcomes had a wide range for assessment tools under visuospatial, attention, problem solving, functional assessment, and mood/daily functioning (see Table 1 for results).

Most of the reviewed tools with cross-language data had similar range of outcomes (both were above 80% or below 80%); however, in 30% of the reviewed assessment tools, the sensitivity and specificity of the tools were above the 80% minimum threshold for English-speakers, and below this rate among Spanish-speakers, although Spanish-language sensitivity and specificity data were not readily available or accessible for all the reviewed tools in this study. The findings of this study indicate that testing and utilization of some of the reviewed instruments with Spanish-speaking individuals might yield lower sensitivity and specificity scores than when used with English-speaking participants.

There was minimal explicit differentiation found between MCI and AD diagnosis in sensitivity reporting. In fact, the TMT-B and FAQ for English-speakers and LASSI-L and ROCF for Spanish-speakers were the only assessments of all differentiated that between MCI and AD diagnosis within sensitivity reporting. Further, the TMT-B was the only assessment in the battery that acknowledged and reported cross-language differentiation. Participant language proved to be a unique finding. Although most assessments engaged native English monolingual speakers, few tests differentiated between participants who were bilingual English and Spanish speakers and native Spanish monolingual speakers. Further, there was no standard reported differentiation in cultural backgrounds within control or assessment groups or consideration of Hispanic and non-Hispanic cultural influences.

The difficulty in locating accurate cross-language electronic data proved to be a limitation in this study, as there were no preexisting standard guidelines for data location and the research team completed continual exhaustive efforts through diverse search terms, databases, and search engines to obtain these scores. Despite this difficulty, the work in this paper represents one of the few papers, to the research team’s knowledge, that provides a critical review of assessment tool gold standard outcomes in one document, which serves to decrease the burden of locating the information contained herein in a myriad of publications and locations for clinicians and researchers. Within assessment tools, the wording and framing of specific questions may contribute to potential bias in gold standard outcomes. The question “which language is your dominant language?” can also be viewed as inherently flawed given that for many bilinguals one language is dominant in one domain whereas a different language is dominant in another domain"(Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012). Additionally, "establishing which language is dominant is to test bilinguals in both languages on an objective measure. However, objective measures can be biased if they are more difficult in one language than the other. Further complicating matters, it is not always clear how to design difficulty-matched measures across different language" (Gollan et al., 2012). In addition, though acculturation is an important consideration in neuropsychological testing, there is limited validity data for the acculturation scales that are available. A full discussion of acculturation measures is outside the scope of the current review, however, Wallace et al. (2010) provided an excellent overview of the available psychometric data on acculturation scales, noting that oftentimes validity was based solely on content validity.

The purpose of this study was to compile and assess a collection of dementia-spectrum assessment tools and evaluate the known sensitivity and specificity of these instruments to aid in uniform reporting. This study attempted to assess the current availability of psychometric data available for neuropsychological instruments, which may be administered to Hispanic and white non-Hispanic older adults presenting for evaluation related to memory disorders. Reviews of this nature are sparse, and this is one of the few to provide an overview of available Spanish-translated neuropsychological instruments and their sensitivity and specificity in detecting dementia-spectrum disorders. This review identified a need for additional neuropsychological instruments to be made available and normed in Spanish to best serve diverse populations, though the development of normative thresholds needs to account for the diversity within the Spanish community. Country and ethnicity-specific testing should be completed to validate an instrument in the precise community within which researchers and clinicians intend administer the measurement tool.

This study integrates gold standard neuropsychological assessment outcomes from the past 50 years (Reitan, 1958) to present. Literature encouraging the investigation of the heterogeneity in psychometric properties and outcomes in the wide range of available assessments is only recently emerging (Costa et al., 2017). The strength of this study lies within the utilization of gold standard criterion for which there can be appropriate discernment for efficacy in assessment utilization. This study should serve as an additional step in this movement toward standardized psychometric properties and outcome reporting in dementia-spectrum assessment within the literature. Recommendations for future cross-language assessments include the acknowledgment and integration of culturally relevant questions within assessments that account for Hispanic cultural, language, and identity nuances, as well as the recognition of gold standard criterion as a method for discernment of a particular assessment's performance ability. This study can act as a bridge between research and clinical practice where researchers and clinicians can find centralized, uniform reporting for appropriate assessment selection and subsequent research engagement and clinical service delivery. This review has demonstrated the need for clinicians to consider their patients' unique cultural and language differences when determining which neuropsychological assessment to utilize for efficiency and efficacy in service delivery. It is recommended that clinicians engaging with culturally diverse populations use sound professional discretion and judgment in their clinical practice when selecting and utilizing a particular neuropsychological assessment in their practice.

Acknowledgments

This study was funded, in part, by the Neuroscience Foundation of Florida, Inc. and the National Institute on Aging (L30 AG060524).

References

Agrell B, & Dehlin O. (1998). The clock-drawing test. Age and Ageing, 27, 399–404. 10.1093/ageing/27.3.399

Altshuler SJ, & Schmautz T. (2006). No Hispanic student left behind: The consequences of “high stakes” testing. Children & Schools, 28, 5–14. 10.1093/cs/28.1.5

Anderson JAE, Saleemi S, & Bialystok E. (2017). Neuropsychological assessments of cognitive aging in monolingual and bilingual older adults. Journal of Neurolinguistics, 43, 17–27. 10.1016/j.jneuroling.2016.08.001 [PubMed: 28392625]

Arango-Lasprilla JC, Rivera D, Aguayo A, Rodríguez W, Garza MT, Saracho CP, . . . Perrin PB (2015). Trail Making Test: Normative data for the Latin American Spanish speaking adult population. NeuroRehabilitation, 37, 639–661. 10.3233/NRE-151284 [PubMed: 26639932]

Ardila A. (2013). The impact of culture on neuropsychological test performance In Uzzell BP, Ponton M, & Ardila A. (Eds.), International handbook of cross-cultural neuropsychology (pp. 23–45). New York, NY: Psychology Press.

Arnold BR, Montgomery GT, Castañeda I, & Longoria R. (1994). Acculturation and performance of Hispanics on selected Halstead-Reitan neuropsychological tests. Assessment, 1, 239–248. 10.1177/107319119400100303

Ashendorf L, Jefferson AL, O’Connor MK, Chaisson C, Green RC, & Stern RA (2008). Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. Archives of Clinical Neuropsychology, 23, 129–137. 10.1016/j.acn.2007.11.005 [PubMed: 18178372]

Barrett JJ, Hart KJ, Schmerler JT, Willmarth K, Carey JA, & Mohammed S. (2009). Criterion validity of the financial skills subscale of the direct assessment of functional status scale. Psychiatry Research, 166, 148–157. 10.1016/j.psychres.2008.01.010 [PubMed: 19269693]

Beier M, Gromisch ES, Hughes AJ, Alschuler KN, Madathil R, Chiaravalloti N, & Foley FW (2017). Proposed cut scores for tests of the Brief International Cognitive Assessment of Multiple Sclerosis (BICAMS). Journal of the Neurological Sciences, 381, 110–116. 10.1016/j.jns.2017.08.019 [PubMed: 28991659]

Bell-McGinty S, Butters MA, Meltzer CC, Greer PJ, Reynolds CF III, & Becker JT (2002). Brain morphometric abnormalities in geriatric depression: Long-term neurobiological effects of illness duration. The American Journal of Psychiatry, 159, 1424–1427. 10.1176/appi.ajp.159.8.1424 [PubMed: 12153839]

Benedict H, Schretlen D, Groninger L, Dobraski M, & Shpritz B. (1996). Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity. Psychological Assessment, 8, 145–153. 10.1037/1040-3590.8.2.145

Benge JF, Balsis S, Geraci L, Massman PJ, & Doody RS (2009). How well do the ADAS-cog and its subscales measure cognitive dysfunction in Alzheimer’s disease? Dementia and Geriatric Cognitive Disorders, 28, 63–69. 10.1159/000230709 [PubMed: 19641319]

Boada M, Cejudo JC, Tàrraga L, López OL, & Kaufer D. (2002). Neuropsychiatric inventory questionnaire (NPI-Q): Spanish validation of an abridged form of the Neuropsychiatric Inventory (NPI). Neurologia, 17, 317–323. [PubMed: 12084358]

Boone KB, Lu P, & Wen J. (2005). Comparison of various RAVLT scores in the detection of noncredible memory performance. Archives of Clinical Neuropsychology, 20, 301–319. 10.1016/j.acn.2004.08.001 [PubMed: 15797167]

Boone KB, Victor TL, Wen J, Razani J, & Pontón M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. Archives of Clinical Neuropsychology, 22, 355–365. 10.1016/j.acn.2007.01.010 [PubMed: 17320344]

Borkowski JG, Benton AL, & Spreen O. (1967). Word fluency and brain damage. Neuropsychologia, 5, 135–140. 10.1016/0028-3932(67)90015-2

Brickman AM, Riba A, Bell K, Marder K, Albert M, Brandt J, & Stern Y. (2002). Longitudinal assessment of patient dependence in Alzheimer disease. Archives of Neurology, 59, 1304–1308. 10.1001/archneur.59.8.1304 [PubMed: 12164728]

Britt DM, Adams SG Jr., Godding PR, Grothues CA, & Varnado P. (1995). Clinical differentiation of the Rey Auditory-Verbal Learning Test. American Journal of Alzheimer's Disease,10,7–18.10.1177/153331759501000603

Burke SL, Rodriguez MJ, Barker W, GreigCusto MT, Rosselli M, Loewenstein DA, & Duara R. (2018). Relationship between cognitive performance and measures of neurodegeneration among Hispanic and white non-Hispanic individuals with normal cognition, mild cognitive impairment, and dementia. Journal of the International Neuropsychological Society, 24, 176–187. [PubMed: 28918757]

Burns A, Lawlor B, & Craig S. (2002). Rating scales in old age psychiatry. The British Journal of Psychiatry, 180, 161–167. 10.1192/bjp.180.2.161 [PubMed: 11823329]

Buschke H, Mowrey WB, Ramratan WS, Zimmerman ME, Loewenstein DA, Katz MJ, & Lipton RB (2017). Memory binding test distinguishes amnestic mild cognitive impairment and dementia from cognitively normal elderly. Archives of Clinical Neuropsychology, 32, 1027–1038. 10.1093/arclin/acx046

Busse M, & Whiteside D. (2012). Detecting suboptimal cognitive effort: Classification accuracy of the Conner's Continuous Performance Test-II, Brief Test Of Attention, and Trail Making Test. The Clinical Neuropsychologist, 26, 675–687. 10.1080/13854046.2012.679623 [PubMed: 22533714]

Cano SJ, Posner HB, Moline ML, Hurt SW, Swartz J, Hsu T, & Hobart JC (2010). The ADAS-cog in Alzheimer's disease clinical trials: Psychometric evaluation of the sum and its parts. Journal of Neurology, Neurosurgery, and Psychiatry, 81, 1363–1368. 10.1136/jnnp.2009.204008

Carleton RN, Thibodeau MA, Teale MJN, Welch PG, Abrams MP, Robinson T, & Asmundson GJG (2013). The center for epidemiologic studies depression scale: A review with a theoretical and empirical examination of item content and factor structure. PLoS ONE, 8, e58067 10.1371/journal.pone.0058067 [PubMed: 23469262]

Cherner M, Suarez P, Lazzaretto D, Fortuny LA, Mindt MR, Dawes S, . . . the HNRC group. (2007). Demographically corrected norms for the Brief Visuospatial Memory Test-revised and Hopkins Verbal Learning Test-revised in monolingual Spanish speakers from the U. S.Mexico border region. Archives of Clinical Neuropsychology, 22, 343–353. 10.1016/j.acn.2007.01.009 [PubMed: 17293078]

Cherner M, Suarez P, Posada C, Fortuny LA, Marcotte T, Grant I, . . . the HNRC group. (2008). Equivalency of Spanish language versions of the trail making test part B including or excluding "CH." The Clinical Neuropsychologist, 22, 662–665. 10.1080/13854040701476976 [PubMed: 17853122]

Chin AL, Negash S, & Hamilton R. (2011). Diversity and disparity in dementia: The impact of ethnoracial differences in Alzheimer disease. Alzheimer Disease and Associated Disorders, 25, 187–195. 10.1097/WAD.0b013e318211c6c9 [PubMed: 21399486]

Connor DJ, & Sabbagh MN (2008). Administration and scoring variance on the ADAS-Cog. Journal of Alzheimer's Disease, 15, 461–464. 10.3233/JAD-2008-15312

Costa A, Bak T, Caffarra P, Caltagirone C, Ceccaldi M, Collette F, . . . Cappa SF (2017). The need for harmonisation and innovation of neuropsychological assessment in neurodegenerative dementias in Europe: Consensus document of the Joint Program for Neurodegenerative Diseases Working Group. Alzheimer's Research & Therapy, 9, 27 10.1186/s13195-017-0254-x

Craft S, Newcomer J, Kanne S, Dagogo-Jack S, Cryer P, Sheline Y, . . . Alderson A. (1996). Memory improvement following induced hyperinsulinemia in Alzheimer's disease. Neurobiology of Aging, 17, 123–130. 10.1016/0197-4580(95)02002-0 [PubMed: 8786794]

Crocco E, Curiel RE, Acevedo A, Czaja SJ, & Loewenstein DA (2014). An evaluation of deficits in semantic cueing and proactive and retroactive interference as early features of Alzheimer's disease. The American Journal of Geriatric Psychiatry, 22, 889–897. 10.1016/j.jagp.2013.01.066 [PubMed: 23768680]

Cummings JL, Mega M, Gray K, RosenbergThompson S, Carusi DA, & Gornbein J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. Neurology, 44, 2308–2308. [PubMed: 7991117]

Davies SR, Field ARJ, Andersen T, & Pestell C. (2011). The ecological validity of the ReyOsterrieth Complex Figure: Predicting everyday problems in children with neuropsychological disorders. Journal of Clinical and Experimental Neuropsychology, 33, 820–831. 10.1080/13803395.2011.574608 [PubMed: 21957867]

Del Ser T, González-Montalvo J-I, MartínezEspinosa S, Delgado-Villapalos C, & Bermejo F. (1997). Estimation of premorbid intelligence in Spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. Brain and Cognition, 33, 343–356. 10.1006/brcg.1997.0877 [PubMed: 9126399]

Demsky YI, Mittenberg W, Quintar B, Katell AD, & Golden CJ (1998). Bias in the use of standard American norms with Spanish translations of the Wechsler Memory Scale-Revised. Assessment, 5, 115–121. 10.1177/107319119800500202 [PubMed: 9626387]

Eaton WW, Smith C, Ybarra M, Muntaner C, & Tien A. (2004). Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESD-R) In Maruish ME (Ed.), The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults (3rd ed., Vol. 3, pp. 363–377). Mahwah, NJ: Erlbaum Publishers.

Echemendia RJ, & Harris JG (2004). Neuropsychological test use with Hispanic/Latino populations in the United States: Part II of a national survey. Applied Neuropsychology, 11, 4–12. 10.1207/s15324826an1101_2 [PubMed: 15471742]

Evans DA, Hebert LE, Beckett LA, Scherr PA, Albert MS, Chown MJ, . . . Taylor JO (1997). Education and other measures of socioeconomic status and risk of incident Alzheimer disease in a defined population of older persons. Archives of Neurology, 54, 1399–1405. 10.1001/archneur.1997.00550230066019 [PubMed: 9362989]

Fernández AL (2013). Development of a confrontation naming test for Spanish-speakers: The Cordoba Naming Test. The Clinical Neuropsychologist, 27, 1179–1198. 10.1080/13854046.2013.822931 [PubMed: 23905612]

Fernández AL, & Fulbright RL (2015). Construct and concurrent validity of the Spanish Adaptation of the Boston Naming Test. Applied Neuropsychology Adult, 22, 355–362. 10.1080/23279095.2014.939178 [PubMed: 25668293]

Fernández-San Martín M, Andrade-Rosa C, Molina JD, Muñoz PE, Carretero B, Rodríguez M, & Silva A. (2002). Validation of the Spanish version of the geriatric depression scale (GDS) in primary care. International Journal of Geriatric Psychiatry, 17, 279–287. 10.1002/gps.588 [PubMed: 11921157]

Ferraro FR (2015). Minority and cross-cultural aspects of neuropsychological assessment: Enduring and emerging trends. London, UK: Psychology Press.

Frank J, & Landeira-Fernandez J. (2008). Comparison between two scoring systems of the Rey-Osterrieth Complex Figure in left and right temporal lobe epileptic patients. Archives of Clinical Neuropsychology, 23, 839–845. 10.1016/j.acn.2008.06.001 [PubMed: 18849141]

García-Herranz S, Díaz-Mardomingo MC, & Peraita H. (2016). Neuropsychological predictors of conversion to probable Alzheimer disease in elderly with mild cognitive impairment. Journal of Neuropsychology, 10, 239–255. 10.1111/jnp.12067 [PubMed: 25809316]

Golden CJ (1978). Stroop color and word test: Cat. no. 30150M; a manual for clinical and experimental uses. Wood Dale, IL: Stoelting.

Gollan TH, Weissberger GH, Runnqvist E, Montoya RI, & Cera CM (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. Bilingualism: Language and Cognition, 15, 594–615. 10.1017/S1366728911000332

González-Palau F, Franco M, Jiménez F, Parra E, Bernate M, & Solis A. (2013). Clinical utility of the Hopkins Verbal Test-Revised for detecting Alzheimer's disease and mild cognitive impairment in

Spanish population. Archives of Clinical Neuropsychology, 28, 245–253. 10.1093/arclin/act004 [PubMed: 23384601]

Gramunt N, Buschke H, Sánchez-Benavides G, Lipton RB, Peña-Casanova J, Diéguez-Vide F, . . . Molinuevo JL (2015). Reference data of the Spanish Memory Binding Test in a midlife population from the ALFA STUDY (Alzheimer's and Family). Journal of Alzheimer's Disease, 48, 613–625. 10.3233/JAD-150237

Gramunt N, Sánchez-Benavides G, Buschke H, Diéguez-Vide F, Peña-Casanova J, Masramon X, . . . Molinuevo JL (2016). The Memory Binding Test: Development of two alternate forms into Spanish and Catalan. Journal of Alzheimer's Disease, 52, 283–293. 10.3233/JAD-151175

Green P, Rohling ML, Lees-Haley PR, & Allen LM III. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. Brain Injury, 15, 1045–1060. 10.1080/02699050110088254 [PubMed: 11712951]

Grober E, & Sliwinsk M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. Journal of Clinical and Experimental Neuropsychology, 13, 933–949. 10.1080/01688639108405109 [PubMed: 1779032]

Gross AL, Mungas DM, Crane PK, Gibbons LE, MacKay-Brandt A, Manly JJ, . . . Jones RN (2015). Effects of education and race on cognitive decline: An integrative study of generalizability versus study-specific results. Psychology and Aging, 30, 863–880. 10.1037/pag0000032 [PubMed: 26523693]

Haan MN, Mungas DM, Gonzalez HM, Ortiz TA, Acharya A, & Jagust WJ (2003). Prevalence of dementia in older Latinos: The influence of type 2 diabetes mellitus, stroke and genetic factors. Journal of the American Geriatrics Society, 51, 169–177. [PubMed: 12558712]

Heun R, Papassotiropoulos A, & Jennssen F. (1998). The validity of psychometric instruments for detection of dementia in the elderly general population. International Journal of Geriatric Psychiatry, 13, 368–380. [PubMed: 9658272]

Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R, . . . the Alzheimer's Disease Neuroimaging Initiative. (2013). Putting the Alzheimer's cognitive test to the test II: Rasch Measurement Theory. Alzheimer's & Dementia, 9, S10–S20. 10.1016/j.jalz.2012.08.006

Hogervorst E, Combrinck M, Lapuerta P, Rue J, Swales K, & Budge M. (2002). The Hopkins Verbal Learning Test and screening for dementia. Dementia and Geriatric Cognitive Disorders, 13, 13–20. 10.1159/000048628 [PubMed: 11731710]

Huff FJ, Collins C, Corkin S, & Rosen TJ (1986). Equivalent forms of the Boston Naming Test. Journal of Clinical and Experimental Neuropsychology, 8, 556–562. 10.1080/01688638608405175 [PubMed: 3805252]

Inman TH, & Berry DT (2002). Cross-validation of indicators of malingering: A comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. Archives of Clinical Neuropsychology, 17, 1–23. 10.1093/arclin/17.1.1 [PubMed: 14589749]

Ivanova I, Salmon DP, & Gollan TH (2013). The multilingual naming test in Alzheimer's disease: Clues to the origin of naming impairments. Journal of the International Neuropsychological Society, 19, 272–283. 10.1017/S1355617712001282 [PubMed: 23298442]

Iverson GL, Lange RT, Green P, & Franzen MD (2002). Detecting exaggeration and malingering with the Trail Making Test. The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D), 16, 398–406. 10.1076/clin.16.3.398.13861

Johnson SC, Silverberg ND, Millis SR, & Hanks RA (2012). Symptom validity indicators embedded in the Controlled Oral Word Association Test. The Clinical Neuropsychologist, 26, 1230–1241. 10.1080/13854046.2012.709886 [PubMed: 22856612]

Juva K, Sulkava R, Erkinjuntti T, Ylikoski R, Valvanne J, & Tilvis R. (1995). Usefulness of the Clinical Dementia Rating scale in screening for dementia. International Psychogeriatrics, 7, 17–24. 10.1017/S1041610295001815 [PubMed: 7579017]

Kaplan E, Goodglass H, Weintraub S, & Goodglass H. (1983). Boston naming test. Philadelphia, PA: Lea & Febiger.

Kolibas E, Korinkova V, Novotny V, Vajdickova K, & Hunakova D. (2000). ADAS-cog (Alzheimer's Disease Assessment Scale-cognitive subscale)—Validation of the Slovak version. Bratislavské Lekárske Listy, 101, 598–602. [PubMed: 11218956]

Lai CK (2014). The merits and problems of Neuropsychiatric Inventory as an assessment tool in people with dementia and other neurological disorders. Clinical Interventions in Aging, 9, 1051–1061. 10.2147/CIA.S63504 [PubMed: 25031530]

Langellier BA, Chen J, Vargas-Bustamante A, Inkelas M, & Ortega AN (2016). Understanding health-care access and utilization disparities among Latino children in the United States. Journal of Child Health Care, 20, 133–144. 10.1177/1367493514555587 [PubMed: 25395597]

Liu OL, Bridgeman B, & Adler RM (2012). Measuring learning outcomes in higher education: Motivation matters. Educational Researcher, 41, 352–362. 10.3102/0013189X12459679

Livney MG, Clark CM, Karlawish JH, Cartmell S, Negrón M, Nuñez J, . . . Arnold SE (2011). Ethnoracial differences in the clinical characteristics of Alzheimer's disease at initial presentation at an urban Alzheimer's disease center. The American Journal of Geriatric Psychiatry, 19, 430–439. 10.1097/JGP.0b013e3181f7d881 [PubMed: 21522051]

Loewenstein DA, Amigo E, Duara R, Guterman A, Hurwitz D, Berkowitz N, . . . Eisdorfer C. (1989). A new scale for the assessment of functional status in Alzheimer's disease and related disorders. Journal of Gerontology, 44, P114–P121. 10.1093/geronj/44.4.P114 [PubMed: 2738312]

Loewenstein DA, Curiel RE, DeKosky S, Rosselli M, Bauer R, Grieg-Custo M, . . . Duara R. (2017). Recovery from proactive semantic interference and MRI volume: A replication and extension study. Journal of Alzheimer's Disease, 59, 131–139. 10.3233/JAD170276

Loewenstein DA, Curiel RE, Greig MT, Bauer RM, Rosado M, Bowers D, . . . Duara R. (2016). A novel cognitive stress test for the detection of preclinical Alzheimer disease: Discriminative properties and relation to amyloid load. The American Journal of Geriatric Psychiatry, 24, 804–813. 10.1016/j.jagp.2016.02.056 [PubMed: 27160985]

Loewenstein D, Curiel RE, Greig-Custo M, Crocco E, Rodriquez R, Barker WW, . . . Duara R. (2015). The relationship between a novel test of semantic interference (LASSI-L) and global and regional accumulation of amyloid in the brains of community-dwelling elders. Alzheimer's & Dementia, 11, P131 10.1016/j.jalz.2015.07.043

Loewenstein DA, Duara R, Argüelles T, & Argüelles S. (1995). Use of the Fuld Object-Memory Evaluation in the detection of mild dementia among Spanish- and English-speaking groups. The American Journal of Geriatric Psychiatry, 3, 300–307. 10.1097/00019442-199503040-00004 [PubMed: 28531063]

Lopez OL, Jagust WJ, Dulberg C, Becker JT, DeKosky ST, Fitzpatrick A, . . . Kuller LH (2003). Risk factors for mild cognitive impairment in the Cardiovascular Health Study Cognition Study: Part 2. Archives of Neurology, 60, 1394–1399. 10.1001/archneur.60.10.1394 [PubMed: 14568809]

López-Villalobos JA, Serrano-Pintado I, Andrés-De Llano JM, Sánchez-Mateos JD, Alberola-López S, & Sánchez-Azón MI (2010). [Usefulness of the Stroop test in attention deficit hyperactivity disorder]. Revista de Neurología, 50, 333–340. [PubMed: 20309831]

Lovejoy DW (1998). Neuropsychological assessment of frontal lobe (executive) functioning in adults with attention deficit hyperactivity disorder [Dissertation]. Norfolk: Virginia Consortium for Professional Psychology.

Low GD, & Hubley AM (2007). Screening for depression after cardiac events using the Beck Depression Inventory-II and the Geriatric Depression Scale. Social Indicators Research, 82, 527–543. 10.1007/s11205-006-9049-3

Lowe DA, & Rogers SA (2011). Estimating premorbid intelligence among older adults: The utility of the AMNART. Journal of Aging Research, 2011, 428132. 10.4061/2011/428132

Mackin RS, Ayalon L, Feliciano L, & Areán PA (2010). The sensitivity and specificity of cognitive screening instruments to detect cognitive impairment in older adults with severe psychiatric illness. Journal of Geriatric Psychiatry and Neurology, 23, 94–99. 10.1177/0891988709358589 [PubMed: 20101070]

Malek-Ahmadi M, Small BJ, & Raj A. (2011). The diagnostic value of controlled oral word association test-FAS and category fluency in single-domain amnestic mild cognitive impairment. Dementia and Geriatric Cognitive Disorders, 32, 235–240. 10.1159/000334525 [PubMed: 22156335]

Manos PJ, & Wu R. (1994). The ten point clock test: A quick screen and grading method for cognitive impairment in medical and surgical patients. International Journal of Psychiatry in Medicine, 24, 229–244. 10.2190/5A0F936P-VG8N-0F5R [PubMed: 7890481]

Marc LG, Raue PJ, & Bruce ML (2008). Screening performance of the 15-item geriatric depression scale in a diverse elderly home care population. The American Journal of Geriatric Psychiatry, 16, 914–921. 10.1097/JGP.0b013e318186bd67 [PubMed: 18978252]

Mast BT, Fitzgerald J, Steinberg J, MacNeill SE, & Lichtenberg PA (2001). Effective screening for Alzheimer's disease among older African Americans. The Clinical Neuropsychologist, 15, 196–202. 10.1076/clin.15.2.196.1892 [PubMed: 11528541]

Matías-Guiu JA, Curiel RE, Rognoni T, Valles-Salgado M, Fernández-Matarrubia M, Hariramani R, . . . Matías-Guiu J. (2017). Validation of the Spanish Version of the LASSI-L for diagnosing mild cognitive impairment and Alzheimer's disease. Journal of Alzheimer's Disease, 56, 733–742. 10.3233/JAD-160866

Mayo A. (2016). Use of the Functional Activities Questionnaire in older adults with dementia. Chicago, IL: Alzheimer's Association Retrieved from https://consultgeri.org/try-this/dementia/d13faq2016r2.pdf

McCabe BE, Vermeesch AL, Hall RF, Peragallo NP, & Mitrani VB (2011). Acculturation and the Center For Epidemiological StudiesDepression Scale for Hispanic women. Nursing Research, 60, 270–275. 10.1097/NNR.0b013e318221b8dc [PubMed: 21677596]

McDougall GJ, Becker H, Vaughan PW, Acee TW, & Delville CL (2010). The revised direct assessment of functional status for independent older adults. The Gerontologist, 50, 363–370. 10.1093/geront/gnp139 [PubMed: 19808842]

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr., Kawas CH, . . . Phelps CH (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia, 7, 263–269. 10.1016/j.jalz.2011.03.005

Mejia A, Hernandez M, Lindsey H, Daughtry J, & Puente A. (2014). A-35Distinctions across Different Spanish WAIS-III Similarities and Digit Span Subtests across Mexican- and Central American-Born Individuals. Archives of Clinical Neuropsychology, 29, 516 10.1093/arclin/acu038.35

Mejia S, Gutiérrez LM, Villa AR, & Ostrosky-Solís F. (2004). Cognition, functional status, education, and the diagnosis of dementia and mild cognitive impairment in Spanish-speaking elderly. Applied Neuropsychology, 11, 194–203. 10.1207/s15324826an1104_4

Mindt MR, Cherner M, Marcotte TD, Moore DJ, Bentley H, Esquivel MM, . . . Group TH (2003). The functional impact of HIV-associated neuropsychological impairment in Spanish-speaking adults: A pilot study. Journal of Clinical and Experimental Neuropsychology, 25, 122–132. [PubMed: 12607177]

Mitrani VB, McCabe BE, Gonzalez-Guarda RM, Florom-Smith A, & Peragallo N. (2013). Participation in SEPA, a sexual and relational health intervention for Hispanic women. Western Journal of Nursing Research, 35, 849–866. 10.1177/0193945913480276 [PubMed: 23493674]

Monllau A, Pena-Casanova J, Blesa R, Aguilar M, Bohm P, Sol JM, & Hernandez G. (2007). Diagnostic value and functional correlations of the ADAS-Cog scale in Alzheimer's disease: Data on NORMACODEM project. Neurologia, 22, 493–501. [PubMed: 17602338]

Monsell SE, Dodge HH, Zhou X-H, Bu Y, Besser LM, Mock C, . . . the Neuropsychology Work Group Advisory to the Clinical Task Force. (2016). Results From the NACC Uniform Data Set Neuropsychological Battery Crosswalk Study. Alzheimer Disease and Associated Disorders, 30, 134–139. 10.1097/WAD.0000000000000111 [PubMed: 26485498]

Morris JC (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. Neurology, 43, 2412–2414. 10.1212/WNL.43.11.2412-a

Mowrey WB, Lipton RB, Katz MJ, Ramratan WS, Loewenstein DA, Zimmerman ME, & Buschke H. (2017). Memory Binding Test predicts incident amnestic mild cognitive impairment. Journal of Alzheimer's Disease, 58, 951–952. 10.3233/JAD-179003

Mungas D, Reed BR, Farias ST, & Decarli C. (2009). Age and education effects on relationships of cognitive test scores with brain structure in demographically diverse older persons. Psychology and Aging, 24, 116–128. 10.1037/a0013421 [PubMed: 19290743]

Neblina C. (2012). Construct and criterion validity of the Rey Auditory Verbal Learning Test-Spanish version in adults with traumatic brain injury. Retrieved from http://digitalscholarship.unlv.edu/thesesdissertations/1686/

O'Bryant SE, Humphreys JD, Schiffer RB, & Sutker PB (2007). Presentation of Mexican Americans to a memory disorder clinic. Journal of Psychopathology and Behavioral Assessment, 29, 137 10.1007/s10862-006-9042-9

O'Bryant SE, Lacritz LH, Hall J, Waring SC, Chan W, Khodr ZG, . . . Cullum CM (2010). Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. Archives of Neurology, 67, 746–749. 10.1001/archneurol.2010.115 [PubMed: 20558394]

O'Bryant SE, Xiao G, Edwards M, Devous M, Gupta VB, Martins R, . . . the Texas Alzheimer's Research and Care Consortium (TARCC). (2013). Biomarkers of Alzheimer's disease among Mexican Americans. Journal of Alzheimer's Disease, 34, 841–849. 10.3233/JAD122074

Oquendo-Jiménez I, Mena R, Antoun MD, & Wojna V. (2010). Linguistic adaptation of the clinical dementia rating scale for a Spanish-speaking population: A focus group approach. Puerto Rico Health Sciences Journal, 29, 102–108. [PubMed: 20496524]

Oscanoa T. (2004). Evaluación de la prueba del reloj en el tamizaje de enfermedad de Alzheimer [Evaluation of the clock test in the screening of Alzheimer's disease]. Anales Universidad Nacional Mayor de San Marcos Facultad de Medicina, 65, 42–48. 10.15381/anales.v65i1.1372

Ostrosky-Solis F, Ramirez M, & Ardila A. (2004). Effects of culture and education on neuropsychological testing: A preliminary study with indigenous and nonindigenous population. Applied Neuropsychology, 11, 186–193.

Parikh RM, Eden DT, Price TR, & Robinson RG (1989). The sensitivity and specificity of the Center for Epidemiologic Studies Depression Scale in screening for post-stroke depression. International Journal of Psychiatry in Medicine, 18, 169–181. 10.2190/BH75-EUYA4FM1-J7QA

Parikh R, Mathai A, Parikh S, Chandra Sekhar G, & Thomas R. (2008). Understanding and using sensitivity, specificity and predictive values. Indian Journal of Ophthalmology, 56, 45–50. 10.4103/0301-4738.37595 [PubMed: 18158403]

Park JH, Jung CH, & Lim JG (1995). The relationships of motor function, education, age and cognitive function to the physical activities of daily living. Journal of Korean Medical Science, 10, 195–199. 10.3346/jkms.1995.10.3.195 [PubMed: 8527046]

Parra MA, Abrahams S, Fabi K, Logie R, Luzzi S, & Della Sala S. (2009). Short-term memory binding deficits in Alzheimer's disease. Brain: A Journal of Neurology, 132, 1057–1066. 10.1093/brain/awp036 [PubMed: 19293236]

Pearson. (2009). Wechsler Memory Scale – Fourth edition. San Antonio, TX: Author Retrieved from http://www.pearsonclinical.com/psychology/products/100000281/wechsler-memory-scale--fourth-ed.wms-iv.html#tab-details

Peña-Casanova J, Monllau A, Böhm P, Aguilar M, Sol JM, Hernández G, . . . the Grupo NORMACODEM. (2005). Diagnostic value and test-retest reliability of the Blessed Dementia Rating Scale for Alzheimer's disease: Data from the NORMACODEM project. Neurologia, 20, 349–355. [PubMed: 16163578]

Pereira FS, Oliveira AM, Diniz BS, Forlenza OV, & Yassuda MS (2010). Cross-cultural adaptation, reliability and validity of the DAFS-R in a sample of Brazilian older adults. Archives of Clinical Neuropsychology, 25, 335–343. 10.1093/arclin/acq029 [PubMed: 20484096]

Perneczky R, Pohl C, Sorg C, Hartmann J, Komossa K, Alexopoulos P, . . . Kurz A. (2006). Complex activities of daily living in mild cognitive impairment: Conceptual and diagnostic issues. Age and Ageing, 35, 240–245. 10.1093/ageing/afj054 [PubMed: 16513677]

Pfeffer RI, Kurosaki TT, Harrah CH Jr., Chance JM, & Filos S. (1982). Measurement of functional activities in older adults in the community. Journal of Gerontology, 37, 323–329. 10.1093/geronj/37.3.323 [PubMed: 7069156]

Podhorna J, Krahnke T, Shear M, Harrison JE, & the Alzheimer's Disease Neuroimaging Initiative. (2016). Alzheimer's Disease Assessment Scale-Cognitive subscale variants in mild cognitive impairment and mild Alzheimer's disease: Change over time and the effect of enrichment strategies. Alzheimer's Research & Therapy, 8, 8 10.1186/s13195-016-0170-5

Powell MR, Locke DE, Smigielski JS, & McCrea M. (2011). Estimating the diagnostic value of the trail making test for suboptimal effort in acquired brain injury rehabilitation patients. The Clinical Neuropsychologist, 25, 108–118. 10.1080/13854046.2010.532912 [PubMed: 21113855]

Ravizza SM, & Carter CS (2008). Shifting set about task switching: Behavioral and neural evidence for distinct forms of cognitive flexibility. Neuropsychologia, 46, 2924–2935. 10.1016/j.neuropsychologia.2008.06.006 [PubMed: 18601941]

Razani J, Burciaga J, Madore M, & Wong J. (2007). Effects of acculturation on tests of attention and information processing in an ethnically diverse group. Archives of Clinical Neuropsychology, 22, 333–341. 10.1016/j.acn.2007.01.008 [PubMed: 17298874]

Razani J, Murcia G, Tabares J, & Wong J. (2007). The effects of culture on WASI test performance in ethnically diverse individuals. The Clinical Neuropsychologist, 21, 776–788. 10.1080/13854040701437481 [PubMed: 17676543]

Reedy SD, Boone KB, Cottingham ME, Glaser DF, Lu PH, Victor TL, . . . Wright MJ (2013). Cross validation of the Lu and colleagues (2003). Rey-Osterrieth Complex Figure Test effort equation in a large known-group sample. Archives of Clinical Neuropsychology, 28, 30–37. 10.1093/arclin/acs106 [PubMed: 23232864]

Reitan RM (1958). Validity of the Trail Making Test as an indicator of organic brain damage. Perceptual and Motor Skills, 8, 271–276. 10.2466/pms.1958.8.3.271

Rohit M, Levine A, Hinkin C, Abramyan S, Saxton E, Valdes-Sueiras M, & Singer E. (2007). Education correction using years in school or reading grade-level equivalent? Comparing the accuracy of two methods in diagnosing HIV-associated neurocognitive impairment. Journal of the International Neuropsychological Society, 13, 462–470. 10.1017/S1355617707070506 [PubMed: 17445295]

Roman F, Iturry M, Rojas G, Barceló E, Buschke H, & Allegri RF (2016). Validation of the Argentine version of the Memory Binding Test (MBT) for early detection of mild cognitive impairment. Dementia & Neuropsychologia, 10, 217–226. 10.1590/S1980-57642016DN1003008 [PubMed: 29213458]

Rosselli M, Ardila A, Santisi MN, Arecco MR, Salvatierra J, Conde A, & Lenis B. (2002). Stroop effect in Spanish-English bilinguals. Journal of the International Neuropsychological Society, 8, 819–827. 10.1017/S1355617702860106 [PubMed: 12240746]

Ruiz-Grosso P, Loret de Mola C, Vega-Dienstmaier JM, Arevalo JM, Chavez K, Vilela A, . . . Huapaya J. (2012). Validation of the Spanish Center for Epidemiological Studies Depression and Zung Self-Rating Depression Scales: A comparative validation study. PLoS ONE, 7, e45413 10.1371/journal.pone.0045413 [PubMed: 23056202]

Sánchez-Benavides G, Manero RM, Quiñones-Ubeda S, de Sola S, Quintana M, & Peña-Casanova J. (2009). Spanish version of the Bayer Activities of Daily Living scale in mild cognitive impairment and mild Alzheimer disease: Discriminant and concurrent validity. Dementia and Geriatric Cognitive Disorders, 27, 572–578. 10.1159/000228259 [PubMed: 19602885]

Sano M, Mackell JA, Ponton M, Ferreira P, Wilson J, Pawluczyk S, . . . Thal LJ (1997). The Spanish Instrument Protocol: Design and implementation of a study to evaluate treatment efficacy Instruments for Spanish-speaking patients with Alzheimer's disease. The Alzheimer's Disease Cooperative Study. Alzheimer Disease and Associated Disorders, 11, S57–S64. 10.1097/00002093-199700112-00009 [PubMed: 9236954]

Sayegh P, & Knight BG (2013). Functional assessment and neuropsychiatric inventory questionnaires: Measurement invariance across Hispanics and non-Hispanic whites. The Gerontologist, 54, 375–386. [PubMed: 23564287]

Schmidt M. (1996). Rey auditory verbal learning test: A handbook. Los Angeles, CA: Western Psychological Services.

Seay JS (2015). Characterizing HIV-associated neurocognitive disorder in two underserved sociodemographic groups [Dissertation]. Miami, FL: University of Miami.

Serrano C, Allegri RF, Drake M, Butman J, Harris P, Nagle C, & Ranalli C. (2001). [A shortened form of the Spanish Boston naming test: A useful tool for the diagnosis of Alzheimer's disease]. Revista de Neurología, 33, 624–627. [PubMed: 11784949]

Siddiqi N, House AO, & Holmes JD (2006). Occurrence and outcome of delirium in medical in-patients: A systematic literature review. Age and Ageing, 35, 350–364. 10.1093/ageing/afl005 [PubMed: 16648149]

Skinner J, Carvalho JO, Potter GG, Thames A, Zelinski E, Crane PK, . . . the Alzheimer's Disease Neuroimaging Initiative. (2012). The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): An expansion of the ADAS-Cog to improve responsiveness in MCI. Brain Imaging and Behavior, 6, 489–501. 10.1007/s11682-012-9166-3 [PubMed: 22614326]

Spedo C. (2014). The Wechsler Memory Scale–fourth ed. (WMS-IV) on memory evaluation of multiple sclerosis patients Retrieved from https://www.academia.edu/23131038/The_Wechsler_Memory_Scale-fourth_edition_WMS-IV_on_memory_evaluation_of_multiple_sclerosis_patients

Spencer RJ, Axelrod BN, Drag LL, Waldron-Perrine B, Pangilinan PH, & Bieliauskas LA (2013). WAIS-IV reliable digit span is no more accurate than age corrected scaled score as an indicator of invalid performance in a veteran sample undergoing evaluation for mTBI. The Clinical Neuropsychologist, 27, 1362–1372. 10.1080/13854046.2013.845248 [PubMed: 24099142]

Suarez PA, Gollan TH, Heaton R, Grant I, Cherner M, & the HNRC Group. (2014). Second-language fluency predicts native language stroop effects: Evidence from Spanish-English bilinguals. Journal of the International Neuropsychological Society, 20, 342–348. 10.1017/S1355617714000058 [PubMed: 24622502]

Sugarman MA, & Axelrod BN (2015). Embed-ded measures of performance validity using verbal fluency tests in a clinical sample. Applied Neuropsychology: Adult, 22, 141–146. 10.1080/23279095.2013.873439 [PubMed: 25153155]

Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, . . . Mayeux R. (2001). Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. Neurology, 56, 49–56. [PubMed: 11148235]

Tappen R, Rosselli M, & Engstrom G. (2010). Evaluation of the Functional Activities Questionnaire (FAQ) in cognitive screening across four American ethnic groups. The Clinical Neuropsychologist, 24, 646–661. [PubMed: 20473827]

Teng E, Becker BW, Woo E, Knopman DS, Cummings JL, & Lu PH (2010). Utility of the functional activities questionnaire for distinguishing mild cognitive impairment from very mild Alzheimer disease. Alzheimer Disease and Associated Disorders, 24, 348–353. 10.1097/WAD.0b013e3181e2fc84 [PubMed: 20592580]

Tuokko H, Hadjistavropoulos T, Miller JA, & Beattie BL (1992). The Clock Test: A sensitive measure to differentiate normal elderly from those with Alzheimer disease. Journal of the American Geriatrics Society, 40, 579–584. 10.1111/j.1532-5415.1992.tb02106.x [PubMed: 1587974]

Umfleet LG, Janecek JK, Quasney E, Sab-sevitz DS, Ryan JJ, Binder JR, & Swanson SJ (2015). Sensitivity and specificity of memory and naming tests for identifying left temporal-lobe epilepsy. Applied Neuropsychology: Adult, 22, 189–196. 10.1080/23279095.2014.895366 [PubMed: 25258176]

Vanderploeg RD (2014). Estimating premorbid level of functioning In Clinician's guide to neuropsychological assessment (pp. 49–78). London, UK: Psychology Press 10.4324/9781410603487

Vega WA, Rodriguez MA, & Gruskin E. (2009). Health disparities in the Latino population. Epidemiologic Reviews, 31, 99–112. 10.1093/epirev/mxp008 [PubMed: 19713270]

Verma N, Beretvas SN, Pascual B, Masdeu JC, Markey MK, & the Alzheimer's Disease Neuroimaging Initiative. (2015). New scoring methodology improves the sensitivity of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) in clinical trials. Alzheimer's Research & Therapy, 7, 64 10.1186/s13195-015-0151-0

Wallace PM, Pomery EA, Latimer AE, Martinez JL, & Salovey P. (2010). A review of acculturation measures and their utility in studies promoting Latino health. Hispanic Journal of Behavioral Sciences, 32, 37–54. [PubMed: 20582238]

Wechsler D. (2004). Wechsler Memory Scale 3, WMS-III manual, Spanish version. San Antonio, TX: Pearson.

Wilkinson GS, & Robertson GJ (2014). Wide Range Achievement Test 4 (WRAT-4). San Antonio, TX: Pearson Retrieved from http://www.pearsonclinical.com/education/products/100001722/wide-range-achievement-test-4--wrat4.html

Wong A, Cheng S-T, Lo ESK, Kwan PWL, Law LSN, Chan AYY, . . . Mok V. (2014). Validity and reliability of the neuropsychiatric inventory questionnaire version in patients with stroke or transient ischemic attack having cognitive impairment. Journal of Geriatric Psychiatry and Neurology, 27, 247–252. 10.1177/0891988714532017 [PubMed: 24763069]

Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, & Leirer VO (1982). Development and validation of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research, 17, 37–49. 10.1016/0022-3956(82)90033-4 [PubMed: 7183759]

Young JC, Caron JE, Baughman BC, & Sawyer RJ (2012). Detection of suboptimal effort with symbol span: Development of a new embedded index. Archives of Clinical Neuropsychology, 27, 159–164. 10.1093/arclin/acr109 [PubMed: 22277126]

Table 1

English and Spanish Sensitivity and Specificity of Instruments Used in the Detection of Cognitive Impairment

Measurement tools	WNH sensitivity	WNH specificity	Spanish sensitivity	Spanish specificity	Domains/Authors
Clinical Dementia Rating Scale (CDR)	71% dementia vs. controls	81% dementia vs. controls	Not available	Not available	Dementia staging CDR Sum of Boxes (SB) English speakers: O'Bryant et al.,2010 Spanish speakers: Oquendo-Jiménez, Mena, Antoun, & Wojna, 2010
Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-Cog)	78% MCI vs. controls	100% MCI vs. controls	95.5% AD vs. controls	72.94% AD vs. controls	Memory, language, praxis, attention, cognition. English speakers: Perneczky et al.,2006 Spanish speakers: Monllau et al.,2007
The Fuld Object Memory Test (FOME or Fuld)	95.5% AD vs. controls	96.7% AD vs. controls	95.9% AD vs. controls	100% AD vs. controls	Memory and learning English speakers: Loewenstein, Duara, Argüelles, & Argüelles, 1995 Spanish speakers: Loewenstein et al., 1995
Short-Term Visual Memory Binding Test	74% amnestic MCI vs. controls	73% amnestic MCI vs. controls	69% amnestic MCI vs. controls	88% amnestic MCI vs. controls	Short-term memory, memory binding English speakers: Buschke et al., 2017 Spanish speakers: Roman et al., 2016
Loewenstein-Acevedo Scales for Semantic Interference and Learning (LASSI-L)	87.9% amnestic MCI vs. controls	89.4%–91.5% amnestic MCI vs. controls	90.1–75.7% amnestic MCI vs. controls	92.9–64.7% amnestic MCI vs. controls	Memory English speakers: Crocco, Curiel, Acevedo, Czaja, & Loewenstein, 2014 Spanish speakers: Matías-Guiu et al., 2017
Brief Visuospatial Memory Test–Revised (BVMT-R)	98% dementia vs. controls	82% dementia vs. controls	Not available	Not available	Visuospatial memory English speakers: Benedict, Schretlen, Groninger, Dobraski, & Shpritz, 1996
Wechsler Memory Scale – fourth Edition (WMS-IV)	96% multiple sclerosis vs. controls	87% multiple sclerosis vs. controls	71%−89%	Not available	Auditory memory; Immediate memory Delayed memory; Visual memory WMS-III English speakers: Spedo, 2014 Spanish speakers: Wechsler, 2004
Wechsler Adult Intelligence Scale (WAIS-IV)	27% for Digit Span subtest of the WAIS Revised	100% for Digit Span subtest of the WAIS Revised	Not available	Not available	Memory English speakers: Inman & Berry, 2002
Hopkins Verbal Learning TestRevised (HVLT-R)	91% AD vs. controls	98% AD vs. controls	96% AD vs. controls	85% AD vs. controls	Memory; Language Function English speakers: Hogervorst et al., 2002 Spanish speakers: Gonzalez-Palau et al., 2013
The Rey Auditory Verbal Learning Test (RAVLT)	73.8% credible vs. noncredible memory	90% credible vs. noncredible memory	65% in distinguishing adverse learning and memory	76% in distinguishing adverse learning and memory	Short-term memory; learning measurement English speakers: Boone, Lu, & Wen, 2005 Spanish speakers: Neblina, 2012
Boston Naming Test (BNT)	57% in predicting left seizure focus	59% in predicting left seizure focus	39% AD vs. controls	89% AD vs. controls	Recalled naming ability English speakers: Umfleet et al., 2015 Spanish speakers: Fernández & Fulbright, 2015
Craft Story 21 Immediate and Delayed	Not available	Not available	Not available	Not available	Ability to remember a short story (Craft et al., 1996)

Measurement tools	WNH sensitivity	WNH specificity	Spanish sensitivity	Spanish specificity	Domains/Authors
Controlled Oral Word Association Test (COWAT)	67% for malingered neurocognitive dysfunction	88% for malingered neurocognitive dysfunction	Not available	Not available	Phonetic and semantic fluency English speakers: Malek-Ahmadi, Small, & Raj, 2011 **See also category fluency**
Multilingual Naming Test (MiNT)	Not available	Not available	80–88% true bilinguals	Not available	Assess naming skills in speakers of multiple languages, detecting naming impairments in monolingual AD and amnestic MCI Spanish speakers: Ivanova, Salmon, & Gollan, 2013
Nelson Adult Reading Test—American English version (AMNART)	83% in distinguishing dementia	81% in distinguishing dementia	Not Applicable	Not Applicable	Premorbid intelligence English speakers: Grober & Sliwinsk, 1991
Wide-Range Achievement Testthird edition-(WRAT-3)	42.9% in distinguishing neuropsychological impairment	88.5% in distinguishing neuropsychological impairment	Not Applicable see WAT	Not Applicable see WAT	Literacy and cognitive reserve—English speakers English speakers: Rohit et al., 2007
Word Accentuation Test (WAT)	Not Applicable See WRAT-3	Not Applicable See WRAT-3	78% mild-moderate dementia vs. controls	82% mild-moderate dementia vs. controls	Premorbid intelligence—Spanish speakers Spanish speakers: Del Ser, González-Montalvo, MartínezEspinosa, Delgado-Villapalos, & Bermejo, 1997
Rey–Osterrieth complex figure (ROCF)	80% credible vs. noncredible	90% credible vs. noncredible	70–85.7% in predicting conversion to dementia after an MCI diagnosis	62–98.3% in predicting conversion to dementia after an MCI diagnosis	Recognition/Recall English speakers: Reedy et al., 2013 Spanish speakers: García-Herranz,Díaz-Mardomingo, & Peraita, 2016
Stroop Color and Word Test (SCWT)	88% in distinguishing cognitive impairment	36.8% in distinguishing cognitive impairment	81% in distinguishing ADHD	72% in distinguishing ADHD	Executive functioning English speakers: Mackin, Ayalon, Feliciano, & Areán, 2010 Spanish speakers: López-Villalobos et al., 2010
Trail Making Test-A (TMT-A)	72% optimal vs. suboptimal effort	82% optimal vs. suboptimal effort	24% in distinguishing HAND	88% in distinguishing HAND	Sequencing, psychomotor, visuomotor speed, spatial tracking, cognitive flexibility, and set shifting ability (executive functioning) English speakers: Powell, Locke, Smigielski, & McCrea, 2011 Spanish speakers: Seay, 2015
Trail Making Test-B (TMT-B)	71.6% MCI/AD vs. controls 69% MCI vs. controls 80.7% AD vs. controls	66.9% MCI/AD vs. controls 66.9% MCI vs. controls 31% AD vs. controls	60% in distinguishing HAND	64% in distinguishing HAND	Problem solving English speakers: Ashendorf et al., 2008 Spanish speakers: Seay, 2015
Clock Face Drawing Test	86% AD vs. controls	92% AD vs. controls	83.9% AD vs. controls	93.5% AD vs. controls	Memory; Executive Function; Language; Attention; Visual Context/Visual-Spatial English speakers: Agrell & Dehlin, 1998 Spanish speakers: Oscanoa, 2004
Functional Activities Questionnaire (FAQ)	80.3% for differentiating AD from MCI	87% for differentiating AD from MCI	Between 36–71% dementia vs. controls	between 79–98% dementia vs. controls	Functional Status Distinguish MCI from AD English speakers: Teng et al., 2010 Spanish speakers: Mejia, Gutiérrez, Villa, & Ostrosky-Solís, 2004
Blessed Dementia Rating Scale (BDRS)	96% in distinguishing dementia	82% in distinguishing dementia	87% in distinguishing AD	90% in distinguishing AD	Memory

Measurement tools	WNH sensitivity	WNH specificity	Spanish sensitivity	Spanish specificity	Domains/Authors
					English speakers: Heun,Papassotiropoulos, & Jennssen, 1998 Spanish speakers: Peña-Casanova et al., 2005
Direct Assessment of Functional Status Scale (DAFS)	75% in distinguishing financial independence	96% in distinguishing financial independence	Not available	Not available	Functional status Financial skills subscale of the DAFS English speakers: Barrett et al., 2009
Geriatric Depression Scale (GDS)	100% in distinguishing major depression	94% in distinguishing major depression	86.7% in distinguishing depression	63.1% in distinguishing depression	Depression in older adults English speakers: Low & Hubley, 2007 Spanish speakers: (Fernández-San Martín et al., 2002)
Center for Epidemiologic Studies Depression Scale (CES-D) Revised	86% in distinguishing post-stroke depression	90% in distinguishing post-stroke depression	91.4% in distinguishing MDE; 77.1% in distinguishing MDE or other psychiatric disorders	96.7% in distinguishing MDE; 79.4% in distinguishing MDE or other psychiatric disorders	Depression symptoms English speakers: Parikh, Eden, Price, & Robinson, 1989 Spanish speakers: Ruiz-Grosso et al., 2012
Neuropsychiatric Inventory Questionnaire (NPI-Q)	74.1% in distinguishing neuropsychiatric symptoms	79.5% in distinguishing neuropsychiatric symptoms	Not available	Not available	Neuropsychiatric Symptoms and severity English speakers: Wong et al., 2014
Marin Acculturation Scale – Short	Not applicable	Not applicable	88%	88%	Acculturation Spanish speakers: Mindt et al., 2003.

Note. AD = Alzheimer disease; HAND = HIV-associated neurocognitive disorders; MCI = mild cognitive impairment; = MDE major depressive episodes.