



Copyright © INS. Published by Cambridge University Press, 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
doi:10.1017/S1355617719000894

CRITICAL REVIEW

A Systematic Review of Neuropsychological Tests for the Assessment of Dementia in Non-Western, Low-Educated or Illiterate Populations

Sanne Franzen¹ , Esther van den Berg¹, Miriam Goudsmit², Caroline K. Jurgens³, Lotte van de Wiel⁴, Yuled Kalkisim¹, Özgül Uysal-Bozkir⁵, Yavuz Ayhan⁶, T. Rune Nielsen⁷ and Janne M. Papma^{1,*} 

¹Department of Neurology, Erasmus MC University Medical Center Rotterdam, the Netherlands

²Department of Medical Psychology, OLVG, Amsterdam, the Netherlands

³Department of Geriatric Medicine, Haaglanden Medical Center, The Hague, the Netherlands

⁴Department of Medical Psychology, Maasstad Ziekenhuis, Rotterdam, the Netherlands

⁵Department of Internal Medicine, Section of Geriatric Medicine, Academic Medical Center, Amsterdam, the Netherlands

⁶Department of Psychiatry, Hacettepe University Faculty of Medicine, Ankara, Turkey

⁷Department of Neurology, Danish Dementia Research Centre, University of Copenhagen, Rigshospitalet, Copenhagen, Denmark

(RECEIVED February 14, 2019; FINAL REVISION June 10, 2019; ACCEPTED July 18, 2019)

Abstract

Objective: Neuropsychological tests are important instruments to determine a cognitive profile, giving insight into the etiology of dementia; however, these tests cannot readily be used in culturally diverse, low-educated populations, due to their dependence upon (Western) culture, education, and literacy. In this review we aim to give an overview of studies investigating domain-specific cognitive tests used to assess dementia in non-Western, low-educated populations. The second aim was to examine the quality of these studies and of the adaptations for culturally, linguistically, and educationally diverse populations. **Method:** A systematic review was performed using six databases, without restrictions on the year or language of publication. **Results:** Forty-four studies were included, stemming mainly from Brazil, Hong Kong, Korea, and considering Hispanics/Latinos residing in the USA. Most studies focused on Alzheimer's disease ($n = 17$) or unspecified dementia ($n = 16$). Memory ($n = 18$) was studied most often, using 14 different tests. The traditional Western tests in the domains of attention ($n = 8$) and construction ($n = 15$), were unsuitable for low-educated patients. There was little variety in instruments measuring executive functioning (two tests, $n = 13$), and language ($n = 12$, of which 10 were naming tests). Many studies did not report a thorough adaptation procedure ($n = 39$) or blinding procedures ($n = 29$). **Conclusions:** Various formats of memory tests seem suitable for low-educated, non-Western populations. Promising tasks in other cognitive domains are the Stick Design Test, Five Digit Test, and verbal fluency test. Further research is needed regarding cross-cultural instruments measuring executive functioning and language in low-educated people. (*JINS*, 2019, 00, 1–21)

Keywords: Alzheimer dementia, Neurodegenerative diseases, Mild cognitive impairment, Cross-cultural comparison, Cognition, Literacy, Education

INTRODUCTION

Over the next decades, a dramatic increase is expected in the number of people living with dementia in developing regions compared to those living in developed regions (Ferri et al., 2005; Prince et al., 2013), due to improvements in life expectancy and rapid population aging, especially in lower- and

middle-income countries (World Health Organization, 2011). In addition, non-Western immigrant populations in Western countries, such as people from Turkey and Morocco who immigrated to Western Europe (Nielsen, Vogel, Phung, Gade, & Waldemar, 2011; Parlevliet et al., 2016), or Hispanic people who immigrated to the USA (Gurland et al., 1997), are reaching an age at which dementia is increasingly prevalent.

Most neuropsychological tests were developed to be used in (educated) Western populations. The work by Howard Andrew Knox in the early 1900s at Ellis Island already

*Correspondence and reprint requests to: Janne M. Papma, Ph.D., Department of Neurology, Erasmus Medical Center, Room Ee-2291, Wytemaweg 80, 3015 CN Rotterdam, the Netherlands. E-mail: j.papma@erasmusmc.nl

showed that adaptations are needed to make tests suitable for populations with diverse backgrounds (Richardson, 2003). It is now widely documented that neuropsychological test performance is substantially affected by factors such as culture, language, (quality of) education, and literacy (Ardila, 2005, 2007; Ardila, Rosselli, & Rosas, 1989; Nielsen & Jorgensen, 2013; Nielsen & Waldemar, 2016; Ostrosky-Solis, Ardila, Rosselli, Lopez-Arango, & Uriel-Mendoza, 1998; Teng, 2002). The rising number of patients with dementia from low-educated and non-Western populations therefore calls for an increase in studies addressing the reliability, validity, and cross-cultural and cross-linguistic applicability of neuropsychological instruments used to assess dementia. Furthermore, these studies should include patients with dementia or mild cognitive impairment (MCI) in their sample to determine whether these tests are sufficiently sensitive and specific to dementia.

Recent studies have mostly focused on developing cognitive screening tests, and an excellent review is available of screening tests that can be used in people who are illiterate (Julayanont & Ruthirago, 2018) and/or low educated (Paddick et al., 2017), as well as reviews about screening tests for specific regions, such as Asia (Rosli, Tan, Gray, Subramanian, & Chin, 2016) and Brazil (Vasconcelos, Brucki, & Bueno, 2007). However, an overview of domain-specific cognitive tests and test batteries that are adapted to or developed for a non-Western, low-educated population is lacking. Domain-specific neuropsychological tests are essential to determine a profile of impaired and intact cognitive functions, providing insights into the underlying etiology of the dementia – something that is not possible with screening tests alone. Furthermore, a comprehensive assessment of the cognitive profile may result in more tailored, personalized care after a diagnosis (Jacova, Kertesz, Blair, Fisk, & Feldman, 2007).

The first aim of this review was to generate an overview of all studies investigating either (1) traditional neuropsychological measures, or adaptations of these measures in non-Western populations with low education levels, or (2) new, assembled neuropsychological tests developed for non-Western, low-educated populations. The second aim was to determine the quality of these studies, and to examine the validity and reliability of the current neuropsychological measures in each cognitive domain, as well as determine which could be applied cross-culturally and cross-linguistically.

METHOD

Identification of Studies

Search terms and databases

Studies were selected based on the title and the abstract. Medline, Embase, Web of Science, Cochrane, Psycinfo, and Google Scholar were used to identify relevant papers, without restrictions on the year of publication or language (for a list of the search terms used, see Supplementary

Material). Studies were included up until August 2018 (no start date). The papers were judged independently by two authors (SF and JMP) according to the inclusion criteria described later. In case of disagreement a consensus agreement was made together with EvdB.

Inclusion criteria

The inclusion criteria were as follows:

1. The study included patients with dementia and/or patients with MCI/Cognitive Impairment No Dementia (CIND).
2. The study was conducted in a non-Western country, or a non-Western population in a Western country. Western was defined as all EU/EEA countries (including Switzerland), Australia, New Zealand, Canada, and the USA. Hispanic/Latino populations in the USA were included in this review as a non-Western population, as this group likely encompasses people with heterogeneous immigration histories and diverse cultural and linguistic backgrounds (Puente & Ardila, 2000).
3. The study described the instrument in sufficient detail for the authors to judge its applicability in a non-Western context, its validity and/or its reliability, that is, it was not merely mentioned as used during a diagnostic/research process, without any further elaboration.

Exclusion criteria

Studies that focused on medical conditions other than dementia were excluded. Screening tests – defined as tests covering multiple domains, but yielding a single total score without individually normed subscores – were also excluded, as some reviews of these already exist (Julayanont & Ruthirago, 2018; Paddick et al., 2017; Rosli et al., 2016; Vasconcelos et al., 2007). Intelligence tests were also excluded from the analysis, except when subtests (e.g. Digit Span) were used to assess dementia in combination with other neuropsychological tests and the study described the cross-cultural applicability. Unpublished dissertations and book chapters were excluded.

Finally, studies that did not include low-educated people were excluded. This was operationalized as studies that did not describe the inclusion of low-educated or illiterate participants in the text, and did not include any education levels lower than primary school in their descriptive tables. An exception was made for studies of which the means and standard deviations of the years of education made it highly likely that low-educated participants were included, defined as a mean number of years of education that did not exceed primary school for the respective country by more than one standard deviation. Data from the UNESCO Institute for Statistics (UNESCO Institute for Statistics, n.d.) were used to determine the length of primary school education for each country.

Data Analysis

Quality assessment

The quality of the studies and the cross-cultural applicability of the instruments was assessed according to eight criteria.

These criteria were developed specifically for this study to reflect important variables in the assessment of low-educated, non-Western persons. Any ambiguous cases with regard to the scoring were resolved in a consensus agreement.

The first criterion was whether any participants who are illiterate were included in the study (“Illiteracy”): 0 = no/not stated, 1 = yes. The second criterion was if the language in which the test was administered was specified (“Language”): 0 = no, 1 = yes. The administration language can significantly influence performance on neuropsychological tests (Boone, Victor, Wen, Razani, & Ponton, 2007; Carstairs, Myers, Shores, & Fogarty, 2006; Kissner, Wendell, Spencer, & Waldstein, 2012), and is especially important in the assessment of immigrants, or in countries where many languages are spoken, such as China (Wong, 2011). Third, the cross-cultural adaptations were scored (“Adaptations”). For this criterion, a modification was made to the system by Beaton, Bombardier, Guillemin, and Ferraz (2000) to capture the aspects relevant to neuropsychological test development: 0 = no procedures mentioned, 1 = translation (and/or back translation) or other changes to the form, but not the concept of the test, such as replacing letters with numbers or colors, 2 = an expert committee reviewed the (back) translation, or stimuli chosen by expert committee, 3 = all of the previous and pretesting, such as a pilot study in healthy controls. Assembled tests were scored either 0, 2, or 3, as no translation and back translation procedures would be required for assembled tests. The fourth criterion was whether the study reported qualitatively on the usefulness of the instrument for clinical practice, such as the acceptability of the material, acceptability of the duration of the test, and/or floor- or ceiling effects (“Feasibility”): 0 = no, 1 = yes. Illiterate people are known to be less test-wise than literate people, potentially affecting the feasibility of a test in this population (Ardila et al., 2010). Fifth, the study was scored on the availability of information on reliability and/or validity: 0 = absent, 1 = either validity or reliability data were described, 2 = both validity and reliability were described. Additionally, three criteria were proposed with regard to the final diagnosis. First, “Circularity” – whether the study described preventive measures against circularity, that is, blinding [similar to the domain “The Reference Standard” in the tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews (Whiting, Rutjes, Reitsma, Bossuyt, & Kleijnen, 2003)]. This was scored: 0 = no/not stated, 1 = yes. Second, “Sources” – whether both neuropsychological and imaging data were used for the diagnosis, and whether a consensus meeting was held: 0 = not specified, 1 = only neuropsychological assessment or imaging, 2 = both neuropsychological assessment and imaging, and (C) for consensus meeting. As misdiagnoses are common in non-Western populations (Nielsen et al., 2011), it is important to rely on multiple sources of data to support the diagnosis. Third, “Criteria” – whether the study reported using subtype-specific dementia criteria: 0 = not specified, 1 = general criteria, such as the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* criteria (American

Psychiatric Association, 1987, 1994, 2000) or the *International Classification of Diseases and Related Health Problems (ICD)* criteria, 2 = extensive clinical criteria, for example, the National Institute on Aging-Alzheimer’s Association (NIA-AA) criteria (McKhann et al., 2011) for Alzheimer’s disease (AD) or the Petersen criteria (Petersen, 2004) for MCI. Although a score of one point on any criterion does not necessarily directly equate with one point on any other criterion, sum scores of these eight quality criteria were calculated for each instrument to provide a general indicator of the quality of the study (with a higher score indicating a higher general quality).

In the following sections and tables, the studies are described by cognitive domain, as defined by cognitive theory and according to standard clinical practice (Lezak, Howieson, Bigler, & Tranel, 2012). Although neuropsychological tests often tap multiple cognitive functions, for example, verbal fluency is a sensitive measure of executive function, but also taps language and memory processes, tests are listed in only one primary cognitive domain. Studies investigating multiple cognitive instruments are described in multiple paragraphs if the tests belong to different cognitive domains. When both Western and non-Western populations are described, only the data for the non-Western group are shown. Discriminative validity is described with the Area Under the Curve (AUC), either for people with dementia *versus* controls or people with MCI *versus* controls (when only people with MCI were included in the study). AUC classification follows the traditional academic point system (<.60 = fail, .60–.69 = poor, .70–.79 = fair, .80–.89 = good, .90–.99 = excellent). When multiple studies reported on the same (partial) study cohort, the study with the most detailed information, the largest study population and/or the most comprehensive dataset is described.

RESULTS

The review process is summarized in Figure 1. The search identified 9869 citations. Furthermore, 23 citations were identified through the reference lists of included studies. After deduplication, 5071 citations remained; these citations were screened on title and abstract. If the topic of the abstract fell within the criteria, but there was insufficient information on the type of population and/or education level that was studied, the participants section and demographic tables in the full text were checked. A total of 81 studies were assessed for eligibility, of which 37 were excluded: 26 due to the fact that low-educated participants were not included in the study sample (see Figure 1).

A total of 44 studies were included in this review. As shown in Figure 2, most studies stemmed from Brazil, the USA (Hispanic/Latino population), Hong Kong, and Korea. Primary school education in these countries lasts 5.46 years on average (with a standard deviation of .74 years and range of 4–7 years). Seventeen studies specifically focused on a population of patients with AD, 16 studies investigated an unspecified dementia group or MCI only,

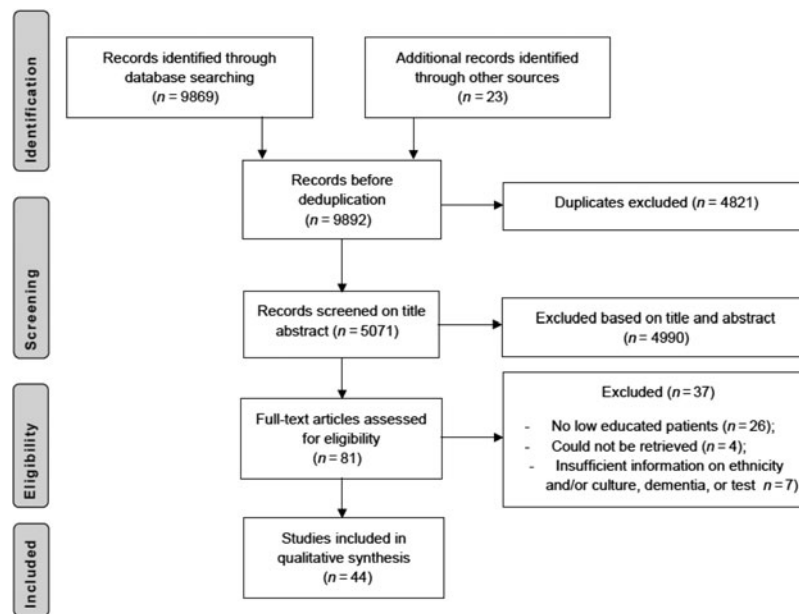


Fig. 1. Results of database searches and selection process.



Fig. 2. Number of studies per country.

and 11 studies investigated a mixed population (mostly AD and smaller groups of other dementias, or AD vs. a “non-AD” group). Of those 11 studies, only one study was specifically aimed at a type of dementia other than AD, that is, Parkinson’s disease dementia (PDD).

Quality criteria scores are summarized in Supplementary Table 1. People who are illiterate were included in 26 of 44 studies. Regarding the tests that were used, 15 studies did not describe performing any translation procedures, and only five studies using an existing test described a complete adaptation procedure with translation, back translation (or other conceptual changes), review by an expert committee, and pretesting (Chan, Tam, Murphy, Chiu, & Lam, 2002; Kim et al., 2017; Lee et al., 2002; Loewenstein, Arguelles, Barker, & Duara, 1993; Shim et al., 2015). The language

the test was administered in, or the fact that it was administered with an interpreter present, was specified in 32 studies. Aspects of the feasibility of the tests were mentioned in 25 studies. With regard to the reference standard, blinding procedures were described in 15 studies. Out of 44 studies, 14 studies made use of both imaging data and neuropsychological assessment to determine the diagnosis, 13 studies used either one of these two and 17 studies did not mention using either imaging data or a neuropsychological assessment to support the final diagnosis. Nearly all studies specified the criteria that were used to determine the diagnosis: the *DSM* or similar criteria were used in 15 studies, and 25 studies used specific clinical criteria. Out of 44 studies, 12 studies reported on both the reliability and the validity of the test.

Attention

Attention tests were described in eight studies, with a total of five different types of tests: the Five Digit Test, the Trail Making Test, the Digit Span subtest of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and WAIS-III, the Corsi Block-Tapping Task, and the WAIS-R Digit Symbol subtest (see Table 1). The Five Digit Test is a relatively new, Stroop-like test, in which participants are asked to either read or count the digits one through five, in congruent and incongruent conditions (e.g. counting two printed fives). With regard to the Trail Making Test, two studies reported on its feasibility. The traditional Trail Making Test could not be used in Chinese and Korean populations with low education levels, leading to “frustration” (Salmon, Jin, Zhang, Grant, & Yu, 1995) and to a 100% failure rate, even in healthy controls (Kim, Baek, & Kim, 2014). An adapted version of Trail Making Test part B, in which participants had to switch between black and white numbers instead of numbers and letters, was completed by a higher percentage of both healthy controls and patients with dementia (Kim et al., 2014). Generally, the AUCs in the domain of attention were variable, ranging from poor to good (.66–.84). In particular, the AUCs for the Digit Span test varied across studies (.69–.84).

Construction and Perception

Construction tests were investigated in 15 studies, by means of five different instruments: the Clock Drawing Test, the Constructional Praxis Test of the neuropsychological test battery of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD), the Stick Design Test, the Block Design subtest of the WAIS-R and of the Wechsler Intelligence Scale for Children-III (WISC-III), and the Object Assembly subtest of the WAIS-R (see Table 2). Of these tests, the Clock Drawing Test was studied most often ($n = 10$). The results with regard to construction tests were mixed. They were described as useful in four studies (Aprahamian, Martinelli, Neri, & Yassuda, 2010; Chan, Yung, & Pan, 2005; Lam et al., 1998; Yap, Ng, Niti, Yeo, & Henderson, 2007), whereas most of the others, such as Salmon et al. (1995), describe this cognitive domain to be “particularly difficult for uneducated subjects” and that some patients “refused to continue because of frustration generated by the difficulty of the task”. The Constructional Praxis Test was evaluated in three studies (Baiyewu et al., 2005; Das et al., 2007; Sahadevan, Lim, Tan, & Chan, 2002), and was compared with the Stick Design Test in one study (Baiyewu et al., 2005). In the Stick Design Test, participants are asked to use matchsticks to copy various printed designs that are similar in complexity to those of the Constructional Praxis Test. The Stick Design Test had lower failure rates (4% vs. 15%) and was also described as “more acceptable” and more sensitive than the Constructional Praxis Test (Baiyewu et al., 2005). Although a study by de Paula, Costa, et al. (2013) also described the Stick Design Test as useful, “eliciting less negative emotional reactions [than

the Constructional Praxis Test] and lowering anxiety levels”, it showed ceiling effects in both healthy controls and patients, similar to the Clock Drawing Test. Generally, the Stick Design Test had fair AUCs of .76 to .79 (Baiyewu et al., 2005; de Paula, Costa, et al., 2013; de Paula, Bertola, et al., 2013). AUCs for the Constructional Praxis were low (Baiyewu et al., 2005), not reported (Das et al., 2007), or left out of the report due to “low diagnostic ability” (Sahadevan et al., 2002). The AUCs were variable for the Clock Drawing Test, ranging from .60 to .87. The Block Design Test had lower sensitivity and specificity in the low educated than high-educated group in one study (Salmon et al., 1995), and different cutoff scores for low and high education levels were recommended in a second study (Sahadevan et al., 2002), as performance was highly influenced by education.

Perception was investigated in two studies, both focusing on olfactory processes. The study by Chan et al. (2002) with the Olfactory Identification Test explicitly describes the adaptation procedure of the test. The authors did a pilot study of 16 odors specific to Hong Kong, and substituted some American items with the items that were most frequently identified as correct in their pilot study. The correct classification rate of the test was 83%. The study by Park et al. (2018) with the Cross-Cultural Smell Identification Test scored positively on only two of the quality criteria and did not provide any sensitivity/specificity data.

Executive Functions

Measures of executive function were investigated in 13 studies (see Table 3), of which 12 studies used the verbal fluency test, mostly focusing on category fluency (i.e. animals, fruits, vegetables). AUCs were fair to excellent for the fluency test (between .79 and .94), although lower sensitivity and specificity were found for lower-educated participants than higher-educated participants in one study (Salmon et al., 1995). Of the six studies that included people who are illiterate (see Table 3), two observed different optimal cutoff scores for illiterate *versus* higher-educated groups (Caramelli, Carthery-Goulart, Porto, Charchat-Fichman, & Nitrini, 2007; Mok, Lam, & Chiu, 2004). Only one study investigated another measure of executive function, the Tower of London test, with low scores for the quality criteria (de Paula et al., 2012). The AUCs for the Tower of London test were good (.80–.90).

Language

Language tests were investigated in 12 studies, with a total of ten tests, or variations thereof (see Table 4). Of these ten tests, only three measured a language function other than naming: the Token Test, the Comprehension subtest of the WAIS-R, and the Vocabulary subtest of the WAIS-R. Information about the discriminative validity was not reported in three studies that used naming tests (Das et al., 2007; Kim et al., 2017; Loewenstein et al., 1993), as well as in all studies using

Table 1. Attention

Author (year)	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
de Paula et al. (2010)	Brazil	Digit Span WAIS-III Corsi Blocks	32	20	17	AD	26	~18	22	70 (5)	~75	71 (6)	5 (4)	~4	4 (4)	Population based	.72 .66	4
de Paula, Bertola, et al. (2013)	Brazil	Digit Span Forward WAIS-III Digit Span Backward WAIS-III	96	93	85	AD	26	21	24	73 (8)	75 (7)	73 (8)	5 (4)	5 (3)	5 (4)	Other	.69 .82	8
de Paula, Querino, Oliveira, Sedo, and Malloy-Diniz (2015)	Brazil	Five Digit Test – Reading Five Digit Test – Counting Five Digit Test – Choosing Five Digit Test – Shifting	40	40	0	–	–	–	–	76 (8) ^a			5 (4) ^a			Outpatient	.72 .75 .70	6
Jacinto et al. (2014)	Brazil	Digit Span Forward (unspecified) Digit Span Backward (unspecified)	202	21	22	–	–	–	–	70 ^b	72 ^b	70 ^b	4 ^b	2 ^b	4 ^b	Outpatient	.69 .72	3
Kim et al. (2014)	Korea	Trail Making Test Black and White	19	11	20	AD	28	20	26	63 (6)	74 (8)	69 (7)	12 (6)	8 (7)	8 (7)	Outpatient	–	6
Loewenstein et al. (1993)	Cuban American (USA)	Digit Span (WAIS-R)	0	38	0	AD	–	16	–	–	72 (6)	–	9 (5)	–	–	Outpatient	–	7
Qiao et al. (2016)	China	Digit Span (WAIS-R)	107 (PD)	33	0	PDD	27	19	–	63 (9)	66 (9)	–	10 (4)	6 (5)	–	Outpatient	.84	4
Salmon et al. (1995) ^c	China	Trail Making Test-A Digit Symbol Substitution (WAIS-R) Digit Span (WAIS-R)	67; 46 ^c	61; 16 ^c	0	–	23; 26 ^c	17; 16 ^c	–	74 (8); 72 (9) ^c	78 (7); 75 (9) ^c	–	–	–	–	Population based	– –	3 3
																	.64–.56; .79–.46 ^c	4

Notes: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; AD = Alzheimer's Dementia; WAIS-R = Wechsler Adult Intelligence Scale-Revised; PDD = Parkinson's Disease Dementia; PD = Parkinson's Disease Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified. – indicates no data available or not applicable.

^a Group total.

^b Median instead of mean.

^c Entire dataset split into uneducated, educated respectively.

Table 2. Construction and perception

Author (year)	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
Aprahamian et al. (2010)	Brazil	Clock Drawing Test	40	66	0	AD	23	15	–	78 (7)	80 (7)	–	0	0	–	Outpatient	.83	8
Baiyewu et al. (2005)	Nigeria	Stick Design Test Constructional Praxis (CERAD)	340	88	296	–	20	11	16	78 (6)	80 (8)	79 (6)	88%	95%	93%	Population based	.78 .69	6
Chan et al. (2002)	Chinese (Hong Kong)	Olfactory Identification Test	12	12	0	AD	27	18	–	74 (6)	76 (5)	–	4 (4)-	5 (6)	–	Outpatient	83% classified correctly	8
Chan et al. (2005)	Chinese (Hong Kong)	Clock Drawing Test	34	51	0	AD, VaD, other	17 ^a			78 (7)	78 (6)	–	3 (3)	3 (3)	–	Outpatient	.81	8
Das et al. (2007)	India	Constructional Praxis (CERAD)	634	0	111	–	29	–	~27	67	–	~68	8 (5)	–	~6	Population based	–	8
de Paula et al. (2010)	Brazil	Clock Drawing Test	32	20	17	AD	26	~18	22	70 (5)	~75	71 (6)	5 (4)	~4	4 (4)	Population based	.79	4
de Paula, Costa, et al. (2013)	Brazil	Stick Design Test Clock Drawing Test	62	93	0	AD	28	21	–	75 ^b	75 ^b	–	4 ^b	4 ^b	–	Outpatient	.76 .84	6
de Paula, Bertola, et al. (2013)	Brazil	Stick Design Test Clock Drawing Test	96	93	85	AD	26	21	24	73 (8)	75 (7)	73 (8)	5 (4)	5 (3)	5 (4)	Other	.77 .87	9
Jacinto et al. (2014)	Brazil	Clock Drawing Test	202	21	22	–	–	–	–	70 ^b	72 ^b	70 ^b	4 ^b	2 ^b	4 ^b	Outpatient	.69–.72	3
Lam et al. (1998)	Chinese (Hong Kong)	Clock Drawing/Reading/Setting	53	53	0	AD, non-AD	–	–	–	74 (7)	77 (9)	–	4 (4) ^a			Mixed sample	.83–.79	6
Loewenstein et al. (1993)	Cuban American (USA)	Block Design (WAIS-R) Object Assembly (WAIS-R)	0	38	0	AD	–	16	–	–	72 (6)	–	–	9 (5)	–	Outpatient	– –	7
Park, Lee, Lee, and Kim (2018)	Korea	Cross-Cultural Smell Identification Test	15	20	78	AD	25 ^a			72 (8) ^a			9 (5) ^a			Outpatient	–	3
Qiao et al. (2016)	China	Block Design (WISC-III)	107			(PD)	33	0	PDD	27	19	–	63 (9)	66 (9)	–	10 (4)	6 (5)	–
Sahadevan et al. (2002)	Chinese (Singapore)	Constructional Praxis (CERAD)	155	72	0	AD	24	16	–	26% ≥ 75	60% ≥ 75	–	54%	67%	–	Outpatient	–	10
Salmon et al. (1995) ^c	China	Block Design (WAIS-R) Object Assembly (WAIS-R) Clock Drawing Test Block Design (WAIS-R)	113	77	0	–	23; 26 ^c	17; 16 ^c	–	74 (8); 72 (9) ^c	78 (7); 75 (9) ^c	–	–	–	–	Population based	.78–.91 .89–.74 .66–.64; .77–.74 ^c	9 9 5 4

(Continued)

Table 2. (Continued)

Author (year)	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
Storey, Rowland, Basic, and Conforti (2002)	Mixed (Australia)	Clock Drawing Test	44	49	0	–	23 ^b	14 ^b	–	76 (8)	80 (7)	–	68%	61%	–	Outpatient	.60–.72	5
Yap et al. (2007)	Chinese (Singapore)	Clock Drawing Test	75	73	0	AD, AD with CVD, VaD	29	17	–	71 (5)	78 (6)	–	8 (5)	5 (4)	–	Outpatient	.84–.85	9

Notes: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; AD = Alzheimer's Dementia; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; VaD = Vascular Dementia; WAIS-R = Wechsler Adult Intelligence Scale-Revised; WISC-III = Wechsler Intelligence Scale for Children-III; PD = Parkinson's Disease; PDD = Parkinson's Disease Dementia; CVD = Cerebrovascular disease.

Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified.

– indicates no data available or not applicable.

^a Group total.

^b Median instead of mean.

^c Entire dataset split into uneducated, educated respectively.

the Comprehension and Vocabulary subtests of the WAIS-R (Loewenstein et al., 1993; Salmon et al., 1995). The AUCs of the Token Test were fair (.76) in both studies (de Paula, Bertola, et al., 2013; de Paula et al., 2010). The naming tests were frequently adapted from the Boston Naming Test, or similar types of tests making use of black-and-white line drawings. The AUCs of the naming tests varied, ranging from poor to excellent (.61–.90), with lower sensitivity and specificity for low educated than high-educated participants in one study (Salmon et al., 1995).

Memory

A total of 14 memory tests were investigated in 18 studies, with stimuli presented to different modalities (visual, auditory, and tactile), and in various formats (cued vs. free recall; word lists vs. stories; see Table 5). Both adaptations of existing tests and some assembled tests were studied, such as a picture-based list learning test from Brazil (Jacinto et al., 2014; Takada et al., 2006) and picture-based cued recall tests in France (Maillet et al., 2016, 2017). AUCs were generally fair to excellent (.74–.99). Remarkably, more than half ($n = 11$) of the studies did not describe blinding procedures (see Table 5). With regard to specific tests, the Fuld Object Memory Evaluation (FOME), using common household objects as stimuli, was used in five studies (Chung, 2009; Loewenstein, Duara, Arguelles, & Arguelles, 1995; Qiao, Wang, Lu, Cao, & Qin, 2016; Rideaux, Beaudreau, Fernandez, & O'Hara, 2012), yielding high sensitivity and specificity rates in most studies, although one found lower sensitivity and specificity in the low-educated group (Salmon et al., 1995). However, the overall quality of the studies investigating this test was relatively low (see Table 5). Tests using a verbal list learning format (Baek, Kim, & Kim, 2012; Chang et al., 2010; de Paula, Bertola, et al., 2013; Sahadevan et al., 2002; Takada et al., 2006) also had good to excellent AUCs (.80–.99). With regard to the modality the stimuli were presented to, one study (Takada et al., 2006) found that a picture-based memory test had better discriminative abilities than a verbal list learning test in the low educated, but not the higher-educated group.

Assessment Batteries

Extensive test batteries were investigated in five studies (see Table 6). The studies by Lee et al. (2002) and Unverzagt et al. (1999) looked into versions of the CERAD neuropsychological test battery. The CERAD battery was specifically designed to create uniformity in assessment methods of AD worldwide (Morris et al., 1989) and contains category verbal fluency (animals), a 15-item version of the Boston Naming Test, the Mini-Mental State Examination, a word list learning task with immediate- and delayed recall, and recognition trials, and the Constructional Praxis Test, including a recall trial. The study by Lee et al. (2002) extensively describes the difficulties in designing an equivalent version

Table 3. Executive functions

Author (year)	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
Aprahamian et al. (2010)	Brazil	CVF – Animals	40	66	0	AD	23	15	–	78 (7)	80 (7)	–	0	0	–	Outpatient	.79	8
Caramelli et al. (2007)	Brazil	CVF– Animals	117	88	0	AD	~25	~18	–	~76	~77	–	~4	~4	–	Outpatient	.91–.81 to .83–1	5
Chiu et al. (1997) ^a	Chinese (Hong Kong)	CVF – Animals CVF – Fruits CVF – Vegetables	53	56	0	AD, VaD, other	27	15	–	74 (7)	77 (9)	–	5 (5)	3 (4)	–	Institutionalized/ Outpatient	.84–.85 .94–.81 .78–.71	7
Das et al. (2007)	India	CVF – Animals CVF – Fruits	634	0	111	–	29	–	~27	67	–	~68	8 (5)	–	~6	Population based	–	8
de Paula et al. (2012)	Brazil	Tower of London	60	60	60	AD	27	21	24	74 (6)	76 (7)	74 (9)	7 (3)	5 (3)	6 (4)	Outpatient	.80–.90	2
de Paula et al. (2010)	Brazil	CVF – Animals	32	20	17	AD	26	~18	22	70 (5)	~75	71 (6)	5 (4)	~4	4 (4)	Population based	.82	4
de Paula, Bertola, et al. (2013)	Brazil	CVF – Animals CVF – Fruits Letter Fluency (S)	96	93	85	AD	26	21	24	73 (8)	75 (7)	73 (8)	5 (4)	5 (3)	5 (4)	Other	.92 .87 .85	8
Jacinto et al. (2014)	Brazil	CVF – Animals	202	21	22	–	–	–	–	70 ^b	72 ^b	70 ^b	4 ^b	2 ^b	4 ^b	Outpatient	.78	3
Loewenstein et al. (1993)	Cuban American (USA)	Letter Fluency (COWAT)	0	38	0	AD	–	16	–	–	72 (6)	–	–	9 (5)	–	Outpatient	–	7
Mok et al. (2004)	Chinese (Hong Kong)	CVF – Animals CVF – Fruits CVF – Vegetables	81	72	0	AD	26	17	–	75 (5)	77 (8)	–	4 (5)	3 (3)	–	Outpatient	.87–.93 to .88–.93	6
Radanovic et al. (2007)	Brazil	CVF – Animals CVF – Fruit	33	24	17	AD, VaD, PDD	23	16	18	77 (5)	79 (5)	77 (7)	2 (3)	2 (3)	0 (1)	Outpatient	.91 .91	7
Sahadevan et al. (2002)	Chinese (Singapore)	CVF – Animals	155	72	0	AD	24	16	–	26% ≥ 75	60% ≥ 75	–	54%	67%	–	Outpatient	.81–.90	9
Salmon et al. (1995) ^c	China	CVF – Animals, Fruits, Vegetables (combined)	113	77	0	–	23; 26 ^c	17; 16 ^c	–	74 (8); 72 (9) ^c	78 (7); 75 (9) ^c	–	–	–	–	Population based	.67–.70; .86–.78 ^c	4

Notes: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; CVF = Category Verbal Fluency; AD = Alzheimer's Dementia; VaD = Vascular Dementia; COWAT = Controlled Oral Word Association Test; PDD = Parkinson's Disease Dementia. Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified.

– indicates no data available or not applicable.

^a Two other fluency categories were described, but not used to assess validity.

^b Median instead of mean.

^c Entire dataset split into uneducated, educated respectively.

Table 4. Language

Author (year)	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
Das et al. (2007)	India	Object Naming Test	634	0	111	–	29	–	~27	67	–	~68	8 (5)	–	~6	Population based	–	8
de Paula et al. (2010)	Brazil	Token Test	32	20	17	AD	26	~18	22	70 (5)	~75	71 (6)	5 (4)	~4	4 (4)	Population based	.76	4
de Paula, Bertola, et al. (2013)	Brazil	TN-LIN Token Test	96	93	85	AD	26	21	24	73 (8)	75 (7)	73 (8)	5 (4)–	5 (3)	5 (4)	Other	.84/.70/.78 .84/.68	8 9
Fernandez (2013)	Argentina	Cordoba Naming Test	26	23	0	AD	–	–	–	74 (7)	76 (9)	–	12 (6)	13 (4)	–	Outpatient	.76	9
Jacinto et al. (2014)	Brazil	Naming Test (BCSB)	202	21	22	–	–	–	–	70 ^a	72 ^a	70 ^a	4 ^a	2 ^a	4 ^a	Outpatient	.61	3
Kim et al. (2017)	Korea	Boston Naming Test-Korean (CERAD)	452	268	0	–	21 (5) ^b	–	–	74 (7) ^b	–	–	6 (5) ^b	–	–	Population based	–	9
Loewenstein et al. (1993)	Cuban American (USA)	Boston Naming Test Comprehension (WAIS-R)	0	38	0	AD	–	16	–	–	72 (6)	–	–	9 (5)	–	Outpatient	–	7
Marquez de la Plata et al. (2008) (also: 2009)	Hispanic (USA)	Texas Spanish Naming Test	55	30	0	–	23	15	–	73 (6)	78 (7)	–	5 ^a	1 ^a	–	Outpatient	.90	5
		Modified Boston Naming Test Spanish															.88	
		15-item Spanish Naming Test															.81	5
Marquez de la Plata et al. (2009)	Colombia	Texas Spanish Naming Test	20	36	0	–	27	17	–	69 (10)	74 (7)	–	9 (4)	6 (5)	–	Outpatient	–	5
		Modified Boston Naming Test Spanish															–	
		Boston Naming Test (CERAD)															–	
Radanovic et al. (2007)	Brazil	Boston Naming Test (CERAD)	33	24	17	AD, VaD, PDD	23	16	18	77 (5)	79 (5)	77 (7)	2 (3)	2 (3)	0 (1)	Outpatient	.76	7
		Naming Test (BCSB)															.88	
Sahadevan et al. (2002)	Chinese (Singapore)	Boston Naming Test	155	72	0	AD	24	16	–	26% ≥ 75	60% ≥ 75	–	54%	67%	–	Outpatient	.63–.83	10
Salmon et al. (1995) ^c	China	Boston Naming Test	113	77	0	–	23; 26 ^c	17; 16 ^c	–	74 (8); 72 (9) ^c	78 (7); 75 (9) ^c	–	–	–	–	Population based	.67–.54; .80–.59 ^c	4
		Vocabulary (WAIS-R)															–	

Notes: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; AD = Alzheimer's Dementia; TN-LIN = The Neuropsychological Investigations Laboratory Naming Test; BCSB = Brief Cognitive Screening Battery; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; WAIS-R = Wechsler Adult Intelligence Scale-Revised; VaD = Vascular Dementia; PDD = Parkinson's Disease Dementia.

Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified.

– indicates no data available or not applicable.

^a Median instead of mean.

^b Group total.

^c Entire dataset split into uneducated, educated respectively.

Table 5. Memory

Author	Population (country)	Instrument	N				MMSE			Age			Education			Research setting	AUC IR or SN-SP	AUC DR or SN-SP	AUC Rec or SN-SP	Quality criteria
			C	D	MCI	Type	C	D	MCI	C	D	MCI	C	D	MCI					
Baek et al. (2012)	Korea	Korean Story Recall Test Seoul Verbal Learning Test	53	72	127	AD	27	23	26	66 (7)	73 (6)	69 (7)	12 (5)	12 (9)	11 (5)	Outpatient	.74	.77	.73	8
Chang et al. (2010)	Taiwan	Chinese Version Verbal Learning Test	217	185	0	AD	29	18	–	71 (10)	79 (7)	–	13 (4)	10 (5)	–	Outpatient	.97	.98	–	9
Chung (2009)	Chinese (Hong Kong)	Fuld Object Memory Evaluation	135	57	0	–	25	16	–	76 (8)	79 (7)	–	33%	47%	–	Outpatient	.97	.93	–	5
Das et al. (2007)	India	Memory (word list)	634	0	111	–	29	–	~27	67	–	~68	8 (5)	–	~6	Population based	–	–	–	10
de Paula, Bertola, et al. (2013)	Brazil	Rey Auditory Verbal Learning Test	96	93	85	AD	26	21	24	73 (8)	75 (7)	73 (8)	5 (4)	5 (3)	5 (4)	Other	.93	.93	.93	8
Grober, Ehrlich, Troche, Hahn, and Lipton (2014)	Latino (USA)	Picture Free and Cued Selective Reminding Test	88	24	0	–	27	24	–	72 (5)	77 (7)	–	8 (4)	6 (4)	–	Population based	.86	–	–	6
Jacinto et al. (2014)	Brazil	List learning (BCSB; picture-based)	202	21	22	–	–	–	–	70 ^a	72 ^a	70 ^a	4 ^a	2 ^a	4 ^a	Outpatient	.76	.80	.80	3
Loewenstein et al. (1993)	Cuban American (USA)	Logical Memory (original WMS) Visual Reproduction (original WMS)	0	38	0	AD	–	16	–	–	72 (6)	–	–	9 (5)	–	Outpatient	–	–	–	7
Loewenstein et al. (1995)	Hispanics (USA)	Fuld Object Memory Evaluation	23	27	0	AD	27	21	–	72 (4)	72 (8)	–	13 (5)	10 (5)	–	Outpatient	.96	–	–	7
Maillet et al. (2017)	Mixed (France)	Memory Associative Test of the district of Seine-Saint-Denis	376	94	0	AD	–	19	–	69 (6)	78 (7)	–	18%	20%	–	Outpatient	.88–.97	–	–	9
Maillet et al. (2016)	Mixed (France)	Test des Neuf Images du 93	282	87	0	–	20 (5) ^b	–	–	70 (7) ^b	–	–	12% ^b	–	–	Outpatient	.87–.96	–	–	7

(Continued)

Table 5. (Continued)

Author	Population (country)	Instrument	N			Type	MMSE			Age			Education			Research setting	AUC IR or SN-SP	AUC DR or SN-SP	AUC Rec or SN-SP	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI					
Qiao et al. (2016)	China	Fuld Object Memory Evaluation	107 (PD)	33	0	PDD	27	19	–	63 (9)	66 (9)	–	10 (4)	6 (5)	–	Outpatient	.73	–	–	4
Rideaux et al. (2012)	Latino (USA)	Fuld Object Memory Evaluation	28	13	27	AD, VaD, other	21 (5) ^b			79 (6) ^b			5 (4) ^b			Population based	.92–.93	–	–	5
Sahadevan et al. (2002)	Chinese (Singapore)	Word list memory	155	72	0	AD	24	16	–	26% ≥ 75	60% ≥ 75	–	54%	67%	–	Outpatient	.87–.82	.93–.91	.85–.84	10
Saka, Mihci, Topcuoglu, and Balkan (2006)	Turkey	Enhanced cued recall	33	62	18	AD versus non-AD	27	18/21	27	73 (7)	74 (6)/ 65 (10)	69 (8)	8 (5)	7 (5)/ 8 (5)	8 (5)	Outpatient	.91	–	–	10
Salmon et al. (1995) ^c	China	Fuld Object Memory Evaluation	113	77	0	–	23; 26	17; 16;	–	74 (8); 72 (9)	78 (7); 75 (9)	–	–	–	–	Population based	.47–.63; .92–.58	–	–	4
Takada et al. (2006)	Brazil	List Learning (CERAD) List learning (BCSB; picture-based)	51	50	0	AD, VaD, PDD, etc.	–	–	–	74 (5); 74 (6)	80 (5); 81 (7)	–	45%	43%	–	Population based	–	.85; .99	–	6
Verghese et al. (2012)	India	Picture Based Memory Impairment Screen	239	65	0	–	27	14	–	67 (6)	72 (7)	–	8 (4)	7 (4)	–	Outpatient	.95–.99	–	–	10

Notes: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; IR = Immediate Recall; SN = Sensitivity at optimal cut-off; SP = Specificity at optimal cut-off; DR = Delayed Recall; Rec = Recognition; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; AD = Alzheimer's Dementia; BCSB = Brief Cognitive Screening Battery; WMS: Wechsler Memory Scale; PD = Parkinson's Disease; PDD = Parkinson's Disease Dementia; VaD = Vascular Dementia; CERAD = Consortium to Establish a Registry for Alzheimer's Disease.

Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified.

– indicates no data available or not applicable.

^a Median instead of mean.

^b Group total.

Entire dataset split into uneducated, educated respectively.

Table 6. Test batteries

Author (year)	Population (country)	Test Battery	N			Type	MMSE			Age			Education			Research setting	AUC (entire battery)	Quality criteria
			C	D	MCI		C	D	MCI	C	D	MCI	C	D	MCI			
Lee et al. (2002)	Korea	CERAD	212	194	–	AD versus non-AD	28	17	–	68 (4)	70 (8)	–	8 (4)	6 (5)	–	Outpatient	–	10
Nielsen et al. (2018)	Mixed (Western Europe)	CNTB	52	41	–	AD versus non-AD	–	–	–	73 (7) ^a	–	–	5 (6) ^a	–	Outpatient	–	8	
Shim et al. (2015)	Korea	LICA	634	0	128	–	26	–	22	72 (6)	–	73 (7)	7 (5)	–	6 (5)	Outpatient	.83	11
Unverzagt et al. (1999)	Jamaica	CERAD	72	20	–	–	23	14	–	79 (6)	82 (6)	–	6 (3)	5 (3)	–	Population based and Outpatient	–	6
Wu et al. (2017)	China	NLCA	50	–	50	–	–	–	–	40% ≥ 65	–	46% ≥ 65	44%	–	52%	Outpatient	.94	7

Subtests of the test batteries

CERAD (Lee et al., 2002)		CNTB (Nielsen et al., 2018)		LICA (Shim et al., 2015)		CERAD (Unverzagt et al., 1999)		NLCA (Wu et al., 2017)	
Subtest	AUC	Subtest	AUC	Subtest	AUC	Subtest	AUC ^b	Subtest	
Boston Naming Test	–	Clock Drawing Test	.79	Digit Stroop	–	Boston Naming Test	42%	Attention subtest	
Constructional Praxis	–	Clock Reading Test	.77	Fluency	–	Boston Naming Test Recall	–	Block Design (“Executive subtest”)	
Constructional Praxis Recall	–	Color Trails Test	~.85	Naming	–	Boston Naming Test Visual Recognition	–	Memory subtest	
Fluency (animals)	–	Copying semi-complex figure	.67	Stick Construction	–	Constructional Praxis	25%	Reasoning subtest	
Word List Memory	–	Copying simple figures	.62	Story Recall	–	Fluency (animals)	58%	Visuospatial function subtest	
		Enhanced cued recall	.96	Visual Recognition	–	Indiana University Token Test	67%		
		Five Digit Test	~.78	Visuospatial Span	–	Word List Memory	83%		
		Fluency (animals)	.90	Word List Memory	–				
		Fluency (supermarket)	.92						
		Picture Naming	.65						
		Recall of Pictures Test	~.93						
		Recall semi-complex figure	.93						

Notes: N = number of participants; MMSE = Mini Mental State Examination; AUC = Area Under the Curve; C = healthy controls; D = dementia; MCI = Mild Cognitive Impairment; CERAD = Consortium to Establish a Registry for Alzheimer’s Disease; AD = Alzheimer’s Dementia; CNTB = European Cross-Cultural Neuropsychological Test Battery; LICA = Literacy Independent Cognitive Assessment; NLCA = Non-Language Based Cognitive Assessment.

Age is mean years (standard deviation); education is presented as mean years (standard deviation) or % low educated or illiterate; MMSE is presented as mean unless otherwise specified.

– indicates no data available or not applicable.

^a Group total.

^b Correct classification rate of dementia patients.

in Korean, most notably with regard to “word frequency, mental imagery, phonemic similarity and semantic or word length equivalence”. In some cases, an adequate translation proved to be “impossible”. Items that used reading and writing (MMSE) were replaced by items concerning judgment to better suit the illiterate population in Korea. The Trail Making Test was added in this study to assess vascular dementia (VaD) and PDD, but – similar to other studies in the domain of attention – less-educated controls had “great difficulties” completing parts A and B of this test. A second study investigated the CERAD in a Jamaican population (Unverzagt et al., 1999). Remarkably, 8 out of 20 dementia patients were “not testable” with the CERAD battery. No further information was supplied as to the cause. The correct classification rates for the patients with dementia that did finish the battery were low (ranging from 25% to 67%) – except for the word list memory test (83%).

A study by Nielsen et al. (2018) investigated the European Cross-Cultural Neuropsychological Test Battery (CNTB) in immigrants with dementia from a Turkish, Moroccan, former Yugoslav, Polish, or Pakistani/Indian background. The CNTB consists of the Rowland Universal Dementia Assessment Scale (RUDAS), the Recall of Pictures Test, Enhanced Cued Recall, the copying and recall of a semi-complex figure, copying of simple figures, the Clock Drawing Test, the Clock Reading Test, a picture naming test, category verbal fluency (animal and supermarket), the Color Trails Test, the Five Digit Test, and serial threes. The Color Trails Test and copy and recall of a semi-complex figure were not administered to participants with less than 1 year of education. The study showed excellent discriminative abilities for measures of memory – Enhanced Cued Recall, Recall of Pictures Test, and recall of a semi-complex figure – and category word fluency. Most of the AUCs for these tests were .90 or higher. Attention measures, that is, the Color Trails Test and Five Digit Test, had fair to good discriminative abilities, with AUCs of around .85 and .78, respectively. The diagnostic accuracy was poor for picture naming (AUC .65) and graphomotor construction tests (AUCs of .62 and .67).

A third battery was the Literacy Independent Cognitive Assessment, or LICA (Shim et al., 2015), a newly developed cognitive battery for people who are illiterate. Subtests include Story and Word Memory, Stick Construction (similar to, but more extensive than the Stick Design Test), a modified Corsi Block Tapping Task, Digit Stroop, category word fluency (animals), a Color and Object Recognition Test, and a naming test. Only the performance on Stick Construction and the Color and Object Recognition Test were not significantly different between controls and MCI patients. The AUC for the entire battery was good (.83) in both the group of people who were literate and the group of people who were illiterate, but no information was provided on the AUCs of the subtests.

The last battery was the Non-Language-based Cognitive Assessment (Wu, Lyu, Liu, Li, & Wang, 2017), a battery primarily designed for aphasia patients, but also validated in Chinese MCI patients. It contains Judgment of Line

Orientation, overlapping figures, a visual reasoning subtest, a visual memory test using stimuli chosen to match the Chinese culture, an attention task in a cross-out paradigm, and Block Design test. All demonstrations were nonverbal. The AUC was excellent (.94), but no information was available regarding the subtests.

DISCUSSION

In this systematic review, an overview was provided of 44 studies investigating domain-specific neuropsychological tests used to assess dementia in non-Western populations with low education levels. The quality of these studies, the reliability, validity, and cross-cultural and/or cross-linguistic applicability were summarized. The studies stemmed mainly from Brazil, Hong Kong, and Korea, or concerned Hispanics/Latinos residing in the USA. Most studies focused on AD or unspecified dementia. Memory was studied most often, and various formats of memory tests seem suitable for low-educated, non-Western populations. The traditional Western tests in the domains of attention and construction were unsuitable for low-educated patients; instead, tests such as the Stick Design Test or Five Digit Test may be considered. There was little variety in instruments measuring executive functioning and language. More cross-cultural studies are needed to advance the assessment of these cognitive domains. With regard to the quality of the studies, the most remarkable findings were that many studies did not report a thorough adaptation procedure or blinding procedures.

A main finding of this review was that most studies investigated either patients with AD or a mixed or unspecified group of patients with dementia or MCI. In practice, this means that it remains unknown whether current domain-specific neuropsychological tests can be used to diagnose other types of dementia in non-Western, low-educated populations. Furthermore, only a third of the included studies described taking procedures against circularity of reasoning, such as blinding, potentially inflating the values for the AUCs. Only a third of the studies made use of both imaging and neuropsychological assessment to determine the reference standard. This can be problematic considering that misdiagnoses are likely to be more prevalent in a population in which barriers to dementia diagnostics in terms of culture, language, and education are present (Daugherty, Puente, Fasfous, Hidalgo-Ruzzante, & Perez-Garcia, 2017; Espino & Lewis, 1998; Nielsen et al., 2011). Another remarkable finding in this review was that only a handful of studies applied a rigorous adaptation procedure in which the instrument was translated, back translated, reviewed by an expert committee, and pilot-tested. These studies highlight the difficulty of developing a test that measures a cognitive construct in the same way as the original test in terms of the language used and the difficulty level. Abou-Mrad et al. (2015) elegantly describe these difficulties and provide details for the interested reader about the way some of these issues were resolved in their study.

With regard to specific cognitive domains, the tests identified in this review that measured attention were the Trail Making Test, WAIS-R Digit Span, Corsi Block Tapping Task, WAIS-R Digit Symbol, and Five Digit Test. It was apparent that traditional Western paper-and-pencil tests (Trail Making Test, Digit Symbol) are hard for uneducated subjects (Kim et al., 2014; Lee et al., 2002; Salmon et al., 1995). It therefore seems unlikely that these types of tests will be useful in low-educated, non-Western populations. With regard to Digit Span tests, previous studies have indicated that performance levels vary depending on the language of administration, for example, due to the way digits are ordered in Spanish *versus* English (Arguelles, Loewenstein, & Arguelles, 2001), or due to a short pronunciation time in Chinese (Stigler, Lee, & Stevenson, 1986). This makes Digit Span less suitable as a measure for cross-linguistic evaluations in diverse populations. On the other hand, the Five Digit Test does not seem to suffer from this limitation: it is described by Sedó (2004) as less influenced by differences in culture, language, and formal education, partially because it only makes use of the numbers one through five, that most illiterate people can identify and use correctly (according to Sedó).

Western instruments used to assess the domain construction, such as the Clock Drawing Test, led to frustration in multiple studies and had limited usefulness in the clinical practice with low-educated patients. This is in line with the finding by Nielsen and Jorgensen (2013), that even healthy illiterate people may experience problems with graphomotor construction tasks. The Stick Design Test, that does not rely on graphomotor responses, was described as more acceptable for low-educated patients. Given the ceiling effects that were present in one study (de Paula, Costa, et al., 2013), as well as the differences in performance between the samples from Nigeria (Baiyewu et al., 2005) and Brazil (de Paula, Costa, et al., 2013), further studies on this instrument are required.

Interestingly, no studies in the domain of Perception and Construction focused specifically on the assessment of visual agnosias, although a test of object recognition and a test with overlapping figures were included in two test batteries. As agnosia is included in the core clinical criteria of probable AD (McKhann et al., 2011), it is important to have the appropriate instruments available to determine whether agnosia is present. The only tests measuring perception were two smell identification tasks (Chan et al., 2002; Park et al., 2018). In recent years, this topic has received more attention from cross-cultural researchers. Although olfactory identification is influenced by experience with specific odors (Ayabe-Kanamura, Saito, Distel, Martinez-Gomez, & Hudson, 1998), and tests would therefore have to be adapted to specific populations, deficits in olfactory perception have been described in the early stages of AD and PDD (Alves, Petrosyan, & Magalhaes, 2014). As this task might also be considered to be ecologically valid, it may be an interesting avenue for further research. The study by Chan et al. (2002) with the Olfactory Identification Test explicitly describes the selection procedure of the scents used in the study, making it easy to adapt to other populations.

With regard to executive functioning, nearly all studies examined the verbal fluency test. In addition, the Tower of London test was examined in one study, and some subtests of attention tests tap aspects of executive functioning as well, such as the incongruent trial of the Five Digit Test or the Color Trails Test part 2. This relative lack of executive functioning tests poses significant problems to the diagnosis of Frontotemporal Dementia (FTD) and other dementias influencing frontal or frontostriatal pathways, such as PDD and dementia with Lewy Bodies (DLB) (Johns et al., 2009; Levy et al., 2002). Although this review shows that a limited amount of research is available on lower-educated populations, studies in higher-educated populations have given some indication of the clinical usefulness of other types of executive functioning tests in non-Western populations. For example, Brazilian researchers (Armentano, Porto, Brucki, & Nitrini, 2009; Armentano, Porto, Nitrini, & Brucki, 2013) found the Rule Shift, Modified Six Elements, and Zoo Map subtests of the Behavioral Assessment of the Dysexecutive Syndrome to be useful in discriminating Brazilian patients with AD from controls. It would be interesting to see whether these subtests can be modified so they can be applied with patients who have little to no formal education.

The results in the cognitive domain of language showed that (adapted) versions of the Boston Naming Test were most often studied. This is remarkable, as it is known that even healthy people who are illiterate are at a disadvantage when naming black-and-white line drawings, such as those in the Boston Naming Test, compared to people who are literate (Reis, Petersson, Castro-Caldas, & Ingvar, 2001). This disadvantage disappears when a test uses colored images or, better yet, real-life objects (Reis, Faisca, Ingvar, & Petersson, 2006; Reis, Petersson, et al., 2001). Considering low-educated patients, Kim et al. (2017) describe an interesting finding: although participants with a low education level scored lower on the naming test, remarkable differential item functioning was discovered; the items “acorn” and “pomegranate” were easier to name for low-educated people than higher-educated people, and the effect was reversed for “compass” and “mermaid”. The authors suggest that this may be due to these groups growing up in rural *versus* urban areas, thereby acquiring knowledge specific to these environments. New naming tests might therefore benefit from differential item functioning analyses with regard to education, but also other demographic variables. It was surprising that none of the studies examined a cross-culturally and cross-linguistically applicable test, even though such a test has been developed, that is, the Cross-Linguistic Naming Test (Ardila, 2007). The Cross-Linguistic Naming Test has been studied in healthy non-Western populations from Morocco, Colombia, and Lebanon (Abou-Mrad et al., 2017; Galvez-Lara et al., 2015), as well as in Spanish patients with dementia (Galvez-Lara et al., 2015). These studies preliminarily support its cross-cultural applicability, although more research is needed in diverse populations with dementia.

Memory was the cognitive domain that was most extensively studied, in different formats and with stimuli presented to different sensory modalities: visual, auditory, and tactile. Both adaptations of existing tests and assembled tests were studied. The memory tests in this review generally had the best discriminative abilities of all cognitive domains that were studied. Although this is a positive finding, given that memory tests play a pivotal role in assessing patients with AD, memory tests alone are insufficient to diagnose, or discriminate between, other types of dementia, such as VaD, DLB, FTD, or PDD.

For the majority of the test batteries that were described, information about the validity of the subtests was not provided. An exception is the study of the CNTB (Nielsen et al., 2018). Largely in line with the other findings in this review, the memory tests of the CNTB performed best, whereas the tests of naming and graphomotor construction performed worst. Attention tests, such as the Color Trails Test and Five Digit Test, performed relatively well. In sum, the CNTB encompasses a variety of potentially useful subtests. Similar to the CNTB, the LICA also includes less traditional tests, such as Stick Construction and Digit Stroop, but the lack of information about the discriminative abilities of the subtests makes it hard to judge the relative value of these tests for the cross-cultural assessment of dementia.

In this review, special attention was paid to the influence of education on the performance on neuropsychological tests. Interestingly, the discriminative abilities of the tests were consistently lower for low-educated participants than high-educated patients (Salmon et al., 1995). It has been suggested that tests with high ecological validity may be more suitable for low-educated populations than the (Western) tests that are currently used. Perhaps inspiration can be drawn from the International Shopping List Test (Thompson et al., 2011) for memory, the Multiple Errands Test for executive functioning (Alderman, Burgess, Knight, & Henman, 2003), or even its Virtual Reality (VR) version (Cipresso et al., 2014), or other VR tests, such as the Non-immersive Virtual Coffee Task (Besnard et al., 2016) or the Multitasking in the City Test (Jovanovski et al., 2012).

Some limitations must be acknowledged with respect to this systematic review. It can be argued that this review should not have been limited to dementia or MCI, and should have also included studies of healthy people – for example, normative data studies – or studies of patients with other medical conditions. The inclusion criterion of patients with dementia or MCI was chosen as it is important to know if and how the presence of dementia influences test performance, before a test can be used in clinical practice. That is: is the test sufficiently sensitive and specific to the presence of disease and to disease progression? If this is not the case, using the test might lead to an underestimation of the presence of dementia, or problems differentiating dementia from other conditions.

Furthermore, with regard to the definition of the target population of this review, questions may be raised whether

African American people from the USA should have been included. Although differences in test performance have indeed been found between African Americans and (non-Hispanic) Whites, these differences mostly appear to be driven by differences in quality of education, as opposed to differences in culture (Manly, Jacobs, Touradji, Small, & Stern, 2002; Nabors, Evans, & Strickland, 2000; Silverberg, Hanks, & Tompkins, 2013). Although a very interesting topic for further research, the absence of cultural or linguistic barriers in this population has led to the exclusion of this population in this review.

Lastly, a remarkable finding was the relative paucity of studies from regions such as Africa and the Middle East. It is important to note that, although the search was thorough and studies in other languages were not excluded from this review, some studies without titles/abstracts in English, or studies that were published in local databases, may not have been found. For example, a review by Fasfous, Al-Joudi, Puente, and Perez-Garcia (2017) describes how Arabic-speaking countries have their own data bases (e.g. Arabpsynet) and how an adequate word for “neuropsychology” is lacking in Arabic. Similar databases are known to exist in other regions as well, such as LILACS in Latin America (Vasconcelos et al., 2007).

A strength of this review is that it provides clinicians and researchers working with non-Western populations with a clear overview of the tests and comprehensive test batteries that may have cross-cultural potential, and could be further studied. For example, researchers might use tests from the CNTB as the basis of the neuropsychological assessment, and supplement it with other tests. If preferred, memory tests can also be chosen from the wide variety of memory tests with good AUCs in this review, such as the Fuld Object Memory Evaluation. Researchers are advised against using measures of attention and construction that are paper-and-pencil based, and instead to use tests such as the Five Digit Test for attention, or the Stick Design Test for construction. With regard to executive functioning, it is recommended to look for new, ecologically valid tests to supplement existing tests such as the category verbal fluency test and the Five Digit Test. Furthermore, it is recommended to use language tests that are not based on black-and-white line drawings, but instead use colored pictures, photographs, or real-life objects. The Cross-Linguistic Naming Test might have potential for such purposes.

Other recommendations for future research are to study patients with a variety of diagnoses, including – but not limited to – FTD, DLB, VaD, and primary progressive aphasia. However, as this review has pointed out, this will remain difficult as long as adequate tests to assess these dementias are lacking. It is therefore recommended that future studies support the diagnosis used as the reference standard by additional biomarkers of disease, such as magnetic resonance imaging scans or lumbar punctures. Another suggestion is to carry out validation studies in patients with dementia for instruments that have only been used in healthy controls or for normative data studies. Lastly, it is recommended that test

developers use the most up-to-date guidelines on the adaptation of cross-cultural tests, such as those by the International Test Commission (International Test Commission, 2017) and others (Hambleton, Merenda, & Spielberger, 2005; Iliescu, 2017), and report in their study how they met the various criteria described in these guidelines.

In conclusion, the neuropsychological assessment of dementia in non-Western, low-educated patients is complicated by a lack of research examining cognitive domains such as executive functioning, non-graphomotor construction, and (the cross-cultural assessment of) language, as well as a lack of studies investigating other types of dementia than AD. However, promising instruments are available in a number of cognitive domains that can be used for future research and clinical practice.

Acknowledgements

This study was supported by grant 733050834 from the Netherlands Organization of Scientific Research (ZonMw Memorabel). The authors would like to thank Wichor Bramer from the Erasmus MC University Medical Center Rotterdam for his help in developing the search strategy.

CONFLICT OF INTEREST

The authors have nothing to disclose.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1355617719000894>.

REFERENCES

- Abou-Mrad, F., Chelune, G., Zamrini, E., Tarabey, L., Hayek, M., & Fadel, P. (2017). Screening for dementia in Arabic: Normative data from an elderly Lebanese sample. *The Clinical Neuropsychologist*, 31(Suppl. 1), 1–19. doi: [10.1080/13854046.2017.1288270](https://doi.org/10.1080/13854046.2017.1288270)
- Abou-Mrad, F., Tarabey, L., Zamrini, E., Pasquier, F., Chelune, G., Fadel, P., & Hayek, M. (2015). Sociolinguistic reflection on neuropsychological assessment: An insight into selected culturally adapted battery of Lebanese Arabic cognitive testing. *Neurological Sciences*, 36(10), 1813–1822. doi: [10.1007/s10072-015-2257-3](https://doi.org/10.1007/s10072-015-2257-3)
- Alderman, N., Burgess, P.W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society: JINS*, 9(1), 31–44.
- Alves, J., Petrosyan, A., & Magalhaes, R. (2014). Olfactory dysfunction in dementia. *World Journal of Clinical Cases*, 2(11), 661–667. doi: [10.12998/wjcc.v2.i11.661](https://doi.org/10.12998/wjcc.v2.i11.661)
- American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.). Washington, DC: American Psychiatric Association Press.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: American Psychiatric Association Press.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th (text revised) ed.). Washington, DC: American Psychiatric Association Press.
- Aprahamian, I., Martinelli, J.E., Neri, A.L., & Yassuda, M.S. (2010). The accuracy of the Clock Drawing Test compared to that of standard screening tests for Alzheimer's disease: Results from a study of Brazilian elderly with heterogeneous educational backgrounds. *International Psychogeriatrics*, 22(1), 64–71. doi: [10.1017/S1041610209991141](https://doi.org/10.1017/S1041610209991141)
- Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review*, 15(4), 185–195. doi: [10.1007/s11065-005-9180-y](https://doi.org/10.1007/s11065-005-9180-y)
- Ardila, A. (2007). Toward the development of a cross-linguistic naming test. *Archives of Clinical Neuropsychology*, 22(3), 297–307. doi: [10.1016/j.acn.2007.01.016](https://doi.org/10.1016/j.acn.2007.01.016)
- Ardila, A., Bertolucci, P.H., Braga, L.W., Castro-Caldas, A., Judd, T., Kosmidis, M.H., & Rosselli, M. (2010). Illiteracy: The neuropsychology of cognition without reading. *Archives of Clinical Neuropsychology*, 25(8), 689–712. doi: [10.1093/arclin/acq07921075867](https://doi.org/10.1093/arclin/acq07921075867)
- Ardila, A., Rosselli, M., & Rosas, P. (1989). Neuropsychological assessment in illiterates: Visuospatial and memory abilities. *Brain and Cognition*, 11(2), 147–166.
- Arguelles, T., Loewenstein, D., & Arguelles, S. (2001). The impact of the native language of Alzheimer's disease and normal elderly individuals on their ability to recall digits. *Aging and Mental Health*, 5(4), 358–365. doi: [10.1080/1360786012008314](https://doi.org/10.1080/1360786012008314)
- Armentano, C.G.D., Porto, C.S., Brucki, S.M.D., & Nitrini, R. (2009). Study on the Behavioural Assessment of the Dysexecutive Syndrome (BADS) performance in healthy individuals, Mild Cognitive Impairment and Alzheimer's disease: A preliminary study. *Dementia & Neuropsychologia*, 3(2), 101–107. doi: [10.1590/S1980-57642009DN30200006](https://doi.org/10.1590/S1980-57642009DN30200006)
- Armentano, C.G.D., Porto, C.S., Nitrini, R., & Brucki, S.M.D. (2013). Ecological evaluation of executive functions in mild cognitive impairment and Alzheimer disease. *Alzheimer Disease & Associated Disorders*, 27(2), 95–101. doi: [10.1097/WAD.0b013e31826540b4](https://doi.org/10.1097/WAD.0b013e31826540b4)
- Ayabe-Kanamura, S., Saito, S., Distel, H., Martinez-Gomez, M., & Hudson, R. (1998). Differences and similarities in the perception of everyday odors. A Japanese-German cross-cultural study. *Annals of the New York Academy of Sciences*, 855, 694–700.
- Baek, M.J., Kim, H.J., & Kim, S. (2012). Comparison between the story recall test and the word-list learning test in Korean patients with mild cognitive impairment and early stage of Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 34(4), 396–404. doi: [10.1080/13803395.2011.645020](https://doi.org/10.1080/13803395.2011.645020)
- Baiyewu, O., Unverzagt, F.W., Lane, K.A., Gureje, O., Ogunniyi, A., Musick, B., & Hendrie, H.C. (2005). The Stick Design test: A new measure of visuoconstructional ability. *Journal of the International Neuropsychological Society: JINS*, 11(5), 598–605. doi: [10.1017/S135561770505071X](https://doi.org/10.1017/S135561770505071X)
- Beaton, D.E., Bombardier, C., Guillemin, F., & Ferraz, M.B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186–3191.
- Besnard, J., Richard, P., Banville, F., Nolin, P., Aubin, G., Le Gall, D., & Allain, P. (2016). Virtual reality and neuropsychological assessment: The reliability of a virtual kitchen to assess daily-life activities in victims of traumatic brain injury. *Applied Neuropsychology-Adult*, 23(3), 223–235. doi: [10.1080/23279095.2015.1048514](https://doi.org/10.1080/23279095.2015.1048514)

- Boone, K.B., Victor, T.L., Wen, J., Razani, J., & Ponton, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology*, 22(3), 355–365. doi: [10.1016/j.acn.2007.01.010](https://doi.org/10.1016/j.acn.2007.01.010)
- Caramelli, P., Carthery-Goulart, M.T., Porto, C.S., Charchat-Fichman, H., & Nitrini, R. (2007). Category fluency as a screening test for Alzheimer disease in illiterate and literate patients. *Alzheimer Disease and Associated Disorders*, 21(1), 65–67. doi: [10.1097/WAD.0b013e31802f244f](https://doi.org/10.1097/WAD.0b013e31802f244f)
- Carstairs, J.R., Myers, B., Shores, E.A., & Fogarty, G. (2006). Influence of language background on tests of cognitive abilities: Australian data. *Australian Psychologist*, 41(1), 48–54. doi: [10.1080/00050060500391878](https://doi.org/10.1080/00050060500391878)
- Chan, A., Tam, J., Murphy, C., Chiu, H., & Lam, L. (2002). Utility of olfactory identification test for diagnosing Chinese patients with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 24(2), 251–259. doi: [10.1076/jcen.24.2.251.992](https://doi.org/10.1076/jcen.24.2.251.992)
- Chan, C.C., Yung, C.Y., & Pan, P.C. (2005). Screening of dementia in Chinese elderly adults by the clock drawing test and the time and change test. *Hong Kong Medical Journal*, 11(1), 13–19.
- Chang, C.C., Kramer, J.H., Lin, K.N., Chang, W.N., Wang, Y.L., Huang, C.W., & Wang, P.N. (2010). Validating the Chinese version of the Verbal Learning Test for screening Alzheimer's disease. *Journal of the International Neuropsychological Society: JINS*, 16(2), 244–251. doi: [10.1017/S1355617709991184](https://doi.org/10.1017/S1355617709991184)
- Chiu, H.F., Chan, C.K., Lam, L.C., Ng, K.O., Li, S.W., Wong, M., & Chan, W.F. (1997). The modified Fuld Verbal Fluency Test: A validation study in Hong Kong. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 52(5), 247–250.
- Chung, J.C. (2009). Clinical validity of Fuld Object Memory Evaluation to screen for dementia in a Chinese society. *International Journal of Geriatric Psychiatry*, 24(2), 156–162. doi: [10.1002/gps.2085](https://doi.org/10.1002/gps.2085)
- Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., & Riva, G. (2014). Virtual multiple errands test (VMET): A virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. *Frontiers in Behavioral Neuroscience*, 8, 405. doi: [10.3389/fnbeh.2014.00405](https://doi.org/10.3389/fnbeh.2014.00405)
- Das, S.K., Bose, P., Biswas, A., Dutt, A., Banerjee, T.K., Hazra, A.M., & Roy, T. (2007). An epidemiologic study of mild cognitive impairment in Kolkata, India. *Neurology*, 68(23), 2019–2026. doi: [10.1212/01.wnl.0000264424.76759.e6](https://doi.org/10.1212/01.wnl.0000264424.76759.e6)
- Daugherty, J.C., Puente, A.E., Fasfous, A.F., Hidalgo-Ruzzante, N., & Perez-Garcia, M. (2017). Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests. *Applied Neuropsychology-Adult*, 24(1), 16–22. doi: [10.1080/23279095.2015.1036992](https://doi.org/10.1080/23279095.2015.1036992)
- de Paula, J.J., Bertola, L., Avila, R.T., Moreira, L., Coutinho, G., de Moraes, E.N., & Malloy-Diniz, L.F. (2013). Clinical applicability and cutoff values for an unstructured neuropsychological assessment protocol for older adults with low formal education. *PLoS One*, 8(9), e73167. doi: [10.1371/journal.pone.0073167](https://doi.org/10.1371/journal.pone.0073167)
- de Paula, J.J., Costa, M.V., Bocardi, M.B., Cortezzi, M., De Moraes, E.N., & Malloy-Diniz, L.F. (2013). The Stick Design Test on the assessment of older adults with low formal education: Evidences of construct, criterion-related and ecological validity. *International Psychogeriatrics*, 25(12), 2057–2065. doi: [10.1017/S1041610213001282](https://doi.org/10.1017/S1041610213001282)
- de Paula, J.J., Moreira, L., Nicolato, R., de Marco, L.A., Correa, H., Romano-Silva, M.A., & Malloy-Diniz, L.F. (2012). The Tower of London Test: Different scoring criteria for diagnosing Alzheimer's disease and mild cognitive impairment. *Psychological Reports*, 110(2), 477–488. doi: [10.2466/03.10.13.PR0.110.2.477-488](https://doi.org/10.2466/03.10.13.PR0.110.2.477-488)
- de Paula, J.J., Querino, E.H., Oliveira, T.D., Sedo, M., & Malloy-Diniz, L.F. (2015). Transcultural issues on the assessment of executive functions and processing speed in older adults with low formal education: Usefulness of the Five Digits Test in the assessment of dementia. *Geriatrics & Gerontology International*, 15(3), 388–389. doi: [10.1111/ggi.12364](https://doi.org/10.1111/ggi.12364)
- de Paula, J.J., Schlottfeldt, C.G., Moreira, L., Cotta, M., Bicalho, M.A., Romano-Silva, M.A., & Malloy-Diniz, L.F. (2010). Psychometric properties of a brief neuropsychological protocol for use in geriatric populations. *Revista de Psiquiatria Clínica*, 37(6), 251–255.
- Espino, D.V. & Lewis, R. (1998). Dementia in older minority populations. Issues of prevalence, diagnosis, and treatment. *The American Journal of Geriatric Psychiatry*, 6(2 Suppl. 1), S19–S25.
- Fasfous, A.F., Al-Joudi, H.F., Puente, A.E., & Perez-Garcia, M. (2017). Neuropsychological measures in the Arab world: A systematic review. *Neuropsychology Review*, 27(2), 158–173. doi: [10.1007/s11065-017-9347-3](https://doi.org/10.1007/s11065-017-9347-3)
- Fernandez, A.L. (2013). Development of a confrontation naming test for Spanish-speakers: The Cordoba Naming Test. *The Clinical Neuropsychologist*, 27(7), 1179–1198. doi: [10.1080/13854046.2013.822931](https://doi.org/10.1080/13854046.2013.822931)
- Ferri, C.P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., & Alzheimer's Disease International (2005). Global prevalence of dementia: A Delphi consensus study. *Lancet*, 366(9503), 2112–2117. doi: [10.1016/S0140-6736\(05\)67889-0](https://doi.org/10.1016/S0140-6736(05)67889-0)
- Galvez-Lara, M., Moriana, J.A., Vilar-Lopez, R., Fasfous, A.F., Hidalgo-Ruzzante, N., & Perez-Garcia, M. (2015). Validation of the cross-linguistic naming test: A naming test for different cultures? A preliminary study in the Spanish population. *Journal of Clinical and Experimental Neuropsychology*, 37(1), 102–112. doi: [10.1080/13803395.2014.1003533](https://doi.org/10.1080/13803395.2014.1003533)
- Grober, E., Ehrlich, A.R., Troche, Y., Hahn, S., & Lipton, R.B. (2014). Screening older Latinos for dementia in the primary care setting. *Journal of the International Neuropsychological Society: JINS*, 20(8), 848–855. doi: [10.1017/S1355617714000708](https://doi.org/10.1017/S1355617714000708)
- Gurland, B., Wilder, D., Lantigua, R., Mayeux, R., Stern, Y., Chen, J., & Killeffer, E. (1997). Differences in rates of dementia between ethnic/racial groups. In L.G. Martin & B.J. Soldo (Eds.), *Racial and Ethnic Differences in the Health of Older Americans* (pp. 233–269). Washington, DC: National Academy Press.
- Hambleton, R.K., Merenda, P.F., & Spielberger, C.D. (Eds.). (2005). *Adapting Educational and Psychological Tests for Cross-cultural Assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Iliescu, D. (2017). *Adapting tests in linguistic and cultural situations*. New York, NY: Cambridge University Press.
- International Test Commission (2017). The ITC guidelines for translating and adapting tests. Retrieved from www.InTestCom.org
- Jacinto, A.F., Brucki, S.M.D., Porto, C.S., de Arruda Martins, M., de Albuquerque Citero, V., & Nitrini, R. (2014). Suggested instruments for general practitioners in countries with low schooling to screen for cognitive impairment in the elderly. *International Psychogeriatrics*, 26(7), 1121–1125. doi: [10.1017/S1041610214000325](https://doi.org/10.1017/S1041610214000325)

- Jacova, C., Kertesz, A., Blair, M., Fisk, J.D., & Feldman, H.H. (2007). Neuropsychological testing and assessment for dementia. *Alzheimers & Dementia*, 3(4), 299–317. doi: [10.1016/j.jalz.2007.07.011](https://doi.org/10.1016/j.jalz.2007.07.011)
- Johns, E.K., Phillips, N.A., Belleville, S., Goupil, D., Babins, L., Kelner, N., & Chertkow, H. (2009). Executive functions in frontotemporal dementia and Lewy body dementia. *Neuropsychology*, 23(6), 765–777. doi: [10.1037/a0016792](https://doi.org/10.1037/a0016792)
- Jovanovski, D., Zakzanis, K., Ruttan, L., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Ecologically valid assessment of executive dysfunction using a novel virtual reality task in patients with acquired brain injury. *Applied Neuropsychology-Adult*, 19(3), 207–220. doi: [10.1080/09084282.2011.643956](https://doi.org/10.1080/09084282.2011.643956)
- Julayanont, P. & Ruthirago, D. (2018). The illiterate brain and the neuropsychological assessment: From the past knowledge to the future new instruments. *Applied Neuropsychology-Adult*, 25(2), 174–187. doi: [10.1080/23279095.2016.1250211](https://doi.org/10.1080/23279095.2016.1250211)
- Kim, B.S., Lee, D.W., Bae, J.N., Kim, J.H., Kim, S., Kim, K.W., & Chang, S.M. (2017). Effects of education on differential item functioning on the 15-item modified Korean version of the Boston Naming Test. *Psychiatry Investigation*, 14(2), 126–135. doi: [10.4306/pi.2017.14.2.126](https://doi.org/10.4306/pi.2017.14.2.126)
- Kim, H.J., Baek, M.J., & Kim, S. (2014). Alternative type of the trail making test in nonnative English-speakers: The trail making test-black & white. *PLoS One*, 9(2), e89078. doi: [10.1371/journal.pone.0089078](https://doi.org/10.1371/journal.pone.0089078)
- Kisser, J.E., Wendell, C.R., Spencer, R.J., & Waldstein, S.R. (2012). Neuropsychological performance of native versus non-native English speakers. *Archives of Clinical Neuropsychology*, 27(7), 749–755. doi: [10.1093/arclin/acs082](https://doi.org/10.1093/arclin/acs082)
- Lam, L.C., Chiu, H.F., Ng, K.O., Chan, C., Chan, W.F., Li, S.W., & Wong, M. (1998). Clock-face drawing, reading and setting tests in the screening of dementia in Chinese elderly adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 53(6), 353–357.
- Lee, J.H., Lee, K.U., Lee, D.Y., Kim, K.W., Jhoo, J.H., Kim, J.H., & Woo, J.I. (2002). Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): Clinical and neuropsychological assessment batteries. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 57(1), 47–53.
- Levy, G., Jacobs, D.M., Tang, M.X., Cote, L.J., Louis, E.D., Alfaró, B., & Marder, K. (2002). Memory and executive function impairment predict dementia in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, 17(6), 1221–1226. doi: [10.1002/mds.10280](https://doi.org/10.1002/mds.10280)
- Lezak, M.D., Howieson, D.B., Bigler, E.D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). New York: Oxford University Press.
- Loewenstein, D.A., Arguelles, T., Barker, W.W., & Duara, R. (1993). A comparative analysis of neuropsychological test performance of Spanish-speaking and English-speaking patients with Alzheimer's disease. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 48(3), 142–149.
- Loewenstein, D.A., Duara, R., Arguelles, T., & Arguelles, 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(4), 300–307. doi: [10.1097/00019442-199503040-00004](https://doi.org/10.1097/00019442-199503040-00004)
- Maillet, D., Matharan, F., Le Clesiau, H., Bailon, O., Peres, K., Amieva, H., & Belin, C. (2016). TNI-93: A new memory test for dementia detection in illiterate and low-educated patients. *Archives of Clinical Neuropsychology*, 31(8), 896–903. doi: [10.1093/arclin/acw065](https://doi.org/10.1093/arclin/acw065)
- Maillet, D., Narme, P., Amieva, H., Matharan, F., Bailon, O., Le Clesiau, H., & Belin, C. (2017). The TMA-93: A new memory test for Alzheimer's disease in illiterate and less educated people. *American Journal of Alzheimer's Disease and Other Dementias*, 32(8), 461–467. doi: [10.1177/1533317517722630](https://doi.org/10.1177/1533317517722630)
- Manly, J.J., Jacobs, D.M., Touradji, P., Small, S.A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society: JINS*, 8(3), 341–348.
- Marquez de la Plata, C., Arango-Lasprilla, J.C., Alegret, M., Moreno, A., Tarraga, L., Lara, M., & Cullum, C.M. (2009). Item analysis of three Spanish naming tests: A cross-cultural investigation. *NeuroRehabilitation*, 24(1), 75–85. doi: [10.3233/NRE-2009-0456](https://doi.org/10.3233/NRE-2009-0456)
- Marquez de la Plata, C., Vicioso, B., Hynan, L., Evans, H.M., Diaz-Arrostia, R., Lacritz, L., & Cullum, C.M. (2008). Development of the Texas Spanish Naming Test: A test for Spanish speakers. *The Clinical Neuropsychologist*, 22(2), 288–304. doi: [10.1080/13854040701250470](https://doi.org/10.1080/13854040701250470)
- McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., & Phelps, C.H. (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. *Alzheimers & Dementia*, 7(3), 263–269. doi: [10.1016/j.jalz.2011.03.005](https://doi.org/10.1016/j.jalz.2011.03.005)
- Mok, E.H., Lam, L.C., & Chiu, H.F. (2004). Category verbal fluency test performance in Chinese elderly with Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 18(2), 120–124. doi: [10.1159/000079190](https://doi.org/10.1159/000079190)
- Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., & Clark, C. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39(9), 1159–1165.
- Nabors, N.A., Evans, J.D., & Strickland, T.L. (2000). Neuropsychological assessment and intervention with African Americans. In E. Fletcher-Janzen, T.L. Strickland, & C.R. Reynolds (Eds.), *Handbook of Cross-Cultural Neuropsychology* (pp. 31–42). New York: Kluwer Academic/Plenum.
- Nielsen, T.R. & Jorgensen, K. (2013). Visuoconstructional abilities in cognitively healthy illiterate Turkish immigrants: A quantitative and qualitative investigation. *The Clinical Neuropsychologist*, 27(4), 681–692. doi: [10.1080/13854046.2013.767379](https://doi.org/10.1080/13854046.2013.767379)
- Nielsen, T.R., Segers, K., Vanderaspolden, V., Beinhoff, U., Minthon, L., Pissioti, A., & Waldemar, G. (2018). Validation of a European Cross-Cultural Neuropsychological Test Battery (CNTB) for evaluation of dementia. *International Journal of Geriatric Psychiatry*, 34(1), 144–152. doi: [10.1002/gps.5002](https://doi.org/10.1002/gps.5002)
- Nielsen, T.R., Vogel, A., Phung, T.K., Gade, A., & Waldemar, G. (2011). Over- and under-diagnosis of dementia in ethnic minorities: A nationwide register-based study. *International Journal of Geriatric Psychiatry*, 26(11), 1128–1135. doi: [10.1002/gps.2650](https://doi.org/10.1002/gps.2650)

- Nielsen, T.R., Vogel, A., Riepe, M.W., de Mendonca, A., Rodriguez, G., Nobili, F., & Waldemar, G. (2011). Assessment of dementia in ethnic minority patients in Europe: A European Alzheimer's Disease Consortium survey. *International Psychogeriatrics*, 23(1), 86–95. doi: [10.1017/S1041610210000955](https://doi.org/10.1017/S1041610210000955)
- Nielsen, T.R. & Waldemar, G. (2016). Effects of literacy on semantic verbal fluency in an immigrant population. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 23(5), 578–590. doi: [10.1080/13825585.2015.1132668](https://doi.org/10.1080/13825585.2015.1132668)
- Ostrosky-Solis, F., Ardila, A., Rosselli, M., Lopez-Arango, G., & Uriel-Mendoza, V. (1998). Neuropsychological test performance in illiterate subjects. *Archives of Clinical Neuropsychology*, 13(7), 645–660.
- Paddick, S.M., Gray, W.K., McGuire, J., Richardson, J., Dotchin, C., & Walker, R.W. (2017). Cognitive screening tools for identification of dementia in illiterate and low-educated older adults, a systematic review and meta-analysis. *International Psychogeriatrics*, 29(6), 897–929. doi: [10.1017/S1041610216001976](https://doi.org/10.1017/S1041610216001976)
- Park, S.J., Lee, J.E., Lee, K.S., & Kim, J.S. (2018). Comparison of odor identification among amnesic and non-amnesic mild cognitive impairment, subjective cognitive decline, and early Alzheimer's dementia. *Neurological Sciences*, 39(3), 557–564. doi: [10.1007/s10072-018-3261-1](https://doi.org/10.1007/s10072-018-3261-1)
- Parlevliet, J.L., Uysal-Bozkir, O., Goudsmit, M., van Campen, J.P., Kok, R.M., Ter Riet, G., & de Rooij, S.E. (2016). Prevalence of mild cognitive impairment and dementia in older non-western immigrants in the Netherlands: A cross-sectional study. *International Journal of Geriatric Psychiatry*, 31(9), 1040–1049. doi: [10.1002/gps.4417](https://doi.org/10.1002/gps.4417)
- Petersen, R.C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3), 183–194. doi: [10.1111/j.1365-2796.2004.01388.x](https://doi.org/10.1111/j.1365-2796.2004.01388.x)
- Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C.P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimers & Dementia*, 9(1), 63–75 e62. doi: [10.1016/j.jalz.2012.11.007](https://doi.org/10.1016/j.jalz.2012.11.007)
- Puente, A.E. & Ardila, A. (2000). Neuropsychological assessment of hispanics. In E. Fletcher-Janzen, T.L. Strickland, & C. Reynolds (Eds.), *Handbook of Cross-Cultural Neuropsychology* (pp. 87–104). New York: Kluwers Academic/Plenum Publishers.
- Qiao, J., Wang, X., Lu, W., Cao, H., & Qin, X. (2016). Validation of neuropsychological tests to screen for dementia in Chinese patients with Parkinson's disease. *American Journal of Alzheimer's Disease and Other Dementias*, 31(4), 368–374. doi: [10.1177/1533317515619478](https://doi.org/10.1177/1533317515619478)
- Radanovic, M., Carthery-Goulart, M.T., Charchat-Fichman, H., Herrera, E., Jr., Lima, E.E.P., Smid, J., & Nitrini, R. (2007). Analysis of brief language tests in the detection of cognitive decline and dementia. *Dementia & Neuropsychologia*, 1(1), 37–45. doi: [10.1590/S1980-57642008DN10100007](https://doi.org/10.1590/S1980-57642008DN10100007)
- Reis, A., Faisca, L., Ingvar, M., & Petersson, K.M. (2006). Color makes a difference: Two-dimensional object naming in literate and illiterate subjects. *Brain and Cognition*, 60(1), 49–54. doi: [10.1016/j.bandc.2005.09.012](https://doi.org/10.1016/j.bandc.2005.09.012)
- Reis, A., Petersson, K.M., Castro-Caldas, A., & Ingvar, M. (2001). Formal schooling influences two- but not three-dimensional naming skills. *Brain and Cognition*, 47(3), 397–411. doi: [10.1006/brcg.2001.1316](https://doi.org/10.1006/brcg.2001.1316)
- Richardson, J.T. (2003). Howard Andrew Knox and the origins of performance testing on Ellis Island, 1912–1916. *History of Psychology*, 6(2), 143–170.
- Rideaux, T., Beaudreau, S.A., Fernandez, S., & O'Hara, R. (2012). Utility of the abbreviated Fuld Object Memory Evaluation and MMSE for detection of dementia and cognitive impairment not dementia in diverse ethnic groups. *Journal of Alzheimer's disease: JAD*, 31(2), 371–386. doi: [10.3233/JAD-2012-112180](https://doi.org/10.3233/JAD-2012-112180)
- Rosli, R., Tan, M.P., Gray, W.K., Subramanian, P., & Chin, A.V. (2016). Cognitive assessment tools in Asia: A systematic review. *International Psychogeriatrics*, 28(2), 189–210. doi: [10.1017/S1041610215001635](https://doi.org/10.1017/S1041610215001635)
- Sahadevan, S., Lim, J.P., Tan, N.J., & Chan, S.P. (2002). Psychometric identification of early Alzheimer disease in an elderly Chinese population with differing educational levels. *Alzheimer Disease and Associated Disorders*, 16(2), 65–72.
- Saka, E., Mihci, E., Topcuoglu, M.A., & Balkan, S. (2006). Enhanced cued recall has a high utility as a screening test in the diagnosis of Alzheimer's disease and mild cognitive impairment in Turkish people. *Archives of Clinical Neuropsychology*, 21(7), 745–751. doi: [10.1016/j.acn.2006.08.007](https://doi.org/10.1016/j.acn.2006.08.007)
- Salmon, D.P., Jin, H., Zhang, M.Y., Grant, I., & Yu, E. (1995). Neuropsychological assessment of Chinese elderly in the Shanghai Dementia Survey. *The Clinical Neuropsychologist*, 9(2), 159–168. doi: [10.1080/13854049508401598](https://doi.org/10.1080/13854049508401598)
- Sedó, M.A. (2004). Test de las cinco cifras: Una alternativa multilingüe y no lectora al test de Stroop ["5 digit test": A multilingual non-reading alternative to the Stroop test]. *Revista de Neurologia*, 38(9), 824–828.
- Shim, Y., Ryu, H.J., Lee, D.W., Lee, J.Y., Jeong, J.H., Choi, S.H., & Ryu, S.H. (2015). Literacy Independent Cognitive Assessment: Assessing Mild Cognitive Impairment in older adults with low literacy skills. *Psychiatry Investigation*, 12(3), 341–348. doi: [10.4306/pi.2015.12.3.341](https://doi.org/10.4306/pi.2015.12.3.341)
- Silverberg, N.D., Hanks, R.A., & Tompkins, S.C. (2013). Education quality, reading recognition, and racial differences in the neuropsychological outcome from traumatic brain injury. *Archives of Clinical Neuropsychology*, 28(5), 485–491. doi: [10.1093/arclin/act023](https://doi.org/10.1093/arclin/act023)
- Stigler, J.W., Lee, S.Y., & Stevenson, H.W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition*, 23(1), 1–20.
- Storey, J.E., Rowland, J.T., Basic, D., & Conforti, D.A. (2002). Accuracy of the clock drawing test for detecting dementia in a multicultural sample of elderly Australian patients. *International Psychogeriatrics*, 14(3), 259–271.
- Takada, L.T., Caramelli, P., Fichman, H.C., Porto, C.S., Bahia, V.S., Anghinah, R., & Nitrini, R. (2006). Comparison between two tests of delayed recall for the diagnosis of dementia. *Arquivos de Neuro-psiquiatria*, 64(1), 35–40.
- Teng, E.L. (2002). Cultural and educational factors in the diagnosis of dementia. *Alzheimer Disease and Associated Disorders*, 16(Suppl. 2), S77–S779.
- Thompson, T.A.C., Wilson, P.H., Snyder, P.J., Pietrzak, R.H., Darby, D., Maruff, P., & Buschke, H. (2011). Sensitivity and test-retest reliability of the International Shopping List Test in assessing verbal learning and memory in mild Alzheimer's disease. *Archives of Clinical Neuropsychology*, 26(5), 412–424. doi: [10.1093/arclin/acr039](https://doi.org/10.1093/arclin/acr039)

- UNESCO Institute for Statistics (n.d.). Primary education, duration (years). All countries and economies. Retrieved from data.uis.unesco.org
- Unverzagt, F.W., Morgan, O.S., Thesiger, C.H., Eldemire, D.A., Luseko, J., Pokuri, S., & Hendrie, H.C. (1999). Clinical utility of CERAD neuropsychological battery in elderly Jamaicans. *Journal of the International Neuropsychological Society: JINS*, 5(3), 255–259.
- Vasconcelos, L.G., Brucki, S.M.D., & Bueno, O.F.A. (2007). Cognitive and functional dementia assessment tools: Review of Brazilian literature. *Dementia & Neuropsychologia*, 1(1), 18–23. doi: [10.1590/S1980-57642008DN10100004](https://doi.org/10.1590/S1980-57642008DN10100004)
- Vergheze, J., Noone, M.L., Johnson, B., Ambrose, A.F., Wang, C., Buschke, H., & Mathuranath, P.S. (2012). Picture-based memory impairment screen for dementia. *Journal of the American Geriatrics Society*, 60(11), 2116–2120. doi: [10.1111/j.1532-5415.2012.04191.x](https://doi.org/10.1111/j.1532-5415.2012.04191.x)
- Whiting, P., Rutjes, A.W., Reitsma, J.B., Bossuyt, P.M., & Kleijnen, J. (2003). The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, 3, 25. doi: [10.1186/1471-2288-3-25](https://doi.org/10.1186/1471-2288-3-25)
- Wong, T. (2011). Neuropsychology of Chinese Americans. In D.E.M. Fujii (Ed.), *Studies on Neuropsychology, Neurology, and Cognition. The Neuropsychology of Asian Americans* (pp. 29–46). New York: Psychology Press.
- World Health Organization (2011). *Global health and aging*. Retrieved from https://www.who.int/ageing/publications/global_health.pdf
- Wu, J.B., Lyu, Z.H., Liu, X.J., Li, H.P., & Wang, Q. (2017). Development and standardization of a new cognitive assessment test battery for Chinese aphasic patients: A preliminary study. *Chinese Medical Journal*, 130(19), 2283–2290. doi: [10.4103/0366-6999.215326](https://doi.org/10.4103/0366-6999.215326)
- Yap, P.L., Ng, T.P., Niti, M., Yeo, D., & Henderson, L. (2007). Diagnostic performance of Clock Drawing Test by CLOX in an Asian Chinese population. *Dementia and Geriatric Cognitive Disorders*, 24(3), 193–200. doi: [10.1159/000107080](https://doi.org/10.1159/000107080)