Received: 18 September 2024 | Revised: 8 February 2025 | Accepted: 29 March 2025

DOI: 10.1002/alz.70207

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

REVIEW ARTICLE

Cognitive assessment tools for Arabic-speaking older adults: A systematic review

Mayssan Kabalan[1] | Ola Bazzi[1] | Josleen Al-Barathie[1] | Ola El Zein[2] | Lara Chehabeddine[1,3] | Joanne Khabsa[3] | Monique Chaaya[1] | Carlos Mendes de Leon[4] | Martine Elbejjani[3,5]

[1]Department of Epidemiology and Population Health, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

[2]Faculty of Medicine, American University of Beirut, Beirut, Lebanon

[3]Clinical Research Institute, Faculty of Medicine, American University of Beirut, Beirut, Lebanon

[4]Department of Global Health, Georgetown University School of Health, Washington, District of Columbia, USA

[5]Department of Internal Medicine, Faculty of Medicine, American University of Beirut, Beirut, Lebanon

Correspondence
Martine Elbejjani, Clinical Research Institute, Department of Internal Medicine, Faculty of Medicine, American University of Beirut, Beirut, Lebanon.
Email: me158@aub.edu.lb

Funding information
National Institutes of Health, Grant/Award Number: 1R01 AG069016-01

Abstract

This systematic review aims to identify available cognitive assessments for Arabic-speaking older adults and to assess their validity and performance. A comprehensive search was conducted using Medline, Embase, and APA PsycInfo up to November 2023, encompassing studies validating or using cognitive tools in Arabic for individuals aged $\geq$ 50. We identified 29 validation studies for 20 cognitive tools and 125 studies using cognitive tools. Three tools were validated in more than one study/setting. Cut-offs for dementia were validated for 16 tools (including two domain-specific tools) and for cognitive impairment for three tools. The Mini-Mental State Examination and Montreal Cognitive Assessment were the most frequently validated and used tools. The results highlight a large need for improved psychometric data for cognitive assessments for Arabic-speaking older adults and identify important gaps in knowledge regarding domain-specific tools, the detection of cognitive changes, and the suitability of assessments across different settings and subgroups.

KEYWORDS
Arabic, cognitive assessment, cognitive tool, dementia, older adults

Highlights

- We reviewed the availability and properties of cognitive assessments in Arabic.
- Psychometric data on cognitive tools for older Arabic-speaking adults are scarce.
- Only three tools are validated in more than one study/setting.
- Data are largely lacking for domain-specific tools and early cognitive changes.
- The review identifies important methodology, reporting, and reproducibility issues.

1 | INTRODUCTION

Alzheimer's disease and related dementias (ADRD) are leading causes of morbidity and impairment in older adults and a global public health concern.[1] The vast majority of available data on ADRD trends, burden, and risk factors come from Western high-income countries, but ≈ 60% of dementia patients live in low- and middle-income countries, where the largest increases in number of cases are expected.[2,3] The

Middle East and North Africa (MENA) region is experiencing an alarming increase in dementia burden, with the prevalence projected to increase by 125% by 2050.[1] Data also show an elevated prevalence of dementia among Arabs living abroad.[4] Despite these trends, ADRD screening, prevention, and risk factor identification efforts remain largely limited in the Arab world, highlighting a critical need for a comprehensive understanding and mapping of existing cognitive assessments and data among Arabic-speaking older adults.

Global research and policy efforts have been prioritizing collecting extensive cognitive data (using different global and domain-specific cognitive tests) in diversified populations to better characterize and track aging- and dementia-related cognitive changes.[5–7] One key example is the Harmonized Cognitive Assessment Protocol (HCAP) project, which has been expanding the administration of cognitive assessments across several populations through the international family of Health and Retirement Studies.[8,9] These efforts recognize the importance of factors such as language, education level, and population-specific cultural and contextual factors for cognitive assessments, their administration, and interpretation.[10,11] They are instrumental to simultaneously advance local ADRD knowledge and strategies and to facilitate cross-population and cross-national data pooling and comparisons. However, cognitive assessment tools have been primarily centered on Western populations, leaving a gap in knowledge about available resources and needs for Arabic-speaking older adults.[12,13]

The Arab world, encompassing 22 countries and a population of 482 million, has several dialect and cultural particularities.[14,15] Arabic speakers use both Modern Standard Arabic (MSA) and diverse colloquial dialects, which vary significantly even within the same country. This linguistic and cultural diversity, combined with socioeconomic differences at both national and individual levels, puts forward further challenges regarding the suitability, transferability, and performance of cognitive tests for a complex and heterogeneous population such as the Arabic-speaking population.[16,17] Importantly, the lack of information on validated tools in Arabic creates major challenges for advancing ADRD research and evidence, and the use of unvalidated and/or unexamined cognitive tests may overlook their transferability and suitability, potentially undermining their performance. A comprehensive review of validation efforts of cognitive assessments for Arabic-speaking older adults is urgently needed to identify gaps, guide tool and cognitive battery development, and harmonize efforts within the Arab world and with global ADRD data and research.

This study systematically reviewed studies that have examined the validity of cognitive assessment tools in Arab-speaking older adult populations (≥ 50 years). The primary objectives were (1) to identify validated cognitive tools in this population and (2) to review validation efforts, including characteristics of validation studies (countries, samples, settings), reported validity methods and data, and performance of validated Arabic cognitive tools. A secondary objective was to review cognitive assessments used in Arabic-speaking populations (whether validated or not) to provide a broader comprehensive insight into existing cognitive data among Arab-speaking older adults.

2 | METHODS

This systematic review was reported in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[18] A protocol for this review was registered on PROSPERO with the number CRD42022379112.

2.1 | Eligibility criteria

For the primary objective (reviewing validity efforts), we included papers reporting on the development and/or validation of cognitive assessment tools in Arabic among people aged ≥ 50. For the secondary objective, we also included studies using any cognitive assessment tool in Arabic language (i.e., using a tool/scale to assess cognitive function whether it was validated or not). We did not restrict eligibility by language of publication or publication date. Development or validation as well as the use of tools in epidemiological and clinical research were considered relevant.

We chose the age limit of 50 years because we aimed to focus on tools validated and used in the context of characterizing cognitive function and impairment in middle-aged and older adults in addition to those validated and used for screening and diagnosis of clinical Alzheimer's disease. Articles were excluded if they were conference abstracts, graduate theses, or reviews and/or if they only used a cognitive assessment to establish inclusion/exclusion criteria (and not to study the cognitive performance of the participants).

2.2 | Information source and search strategy

We searched three databases: Medline (Ovid interface, 1946 onward), Embase (Ovid interface, 1947 onward), and APA PsycInfo (1887 onward) on November 26, 2022. The search was updated toward the end of the review on November 17, 2023, to make sure all relevant studies were included. In addition, the reference lists of included studies were screened. The search strategies for the three databases were developed in close coordination with the head of the medical library at the American University of Beirut (O.E.Z.), an expert in conducting systematic reviews. The search strategy for Medline was developed using medical subject headings (MeSH) and text words related to four main concepts: (1) cognitive function, cognitive status, or dementia; (2) assessment tools; (3) Arabic or Arab countries; and (4) older adults (Appendix SA in supporting information). The Medline strategy was then adapted to conform to the syntax and subject headings of Embase and APA PsycInfo.

2.3 | Screening and study selection

The results of our literature search were uploaded to a reference management software package (Endnote) to remove duplicates. The library was then moved to Rayyan, a web-based tool for managing and screening references in systematic reviews, for screening and further duplicate removal. We developed and pilot-tested two standardized screening forms based on our inclusion and exclusion criteria: a title and abstract screening form and a full-text screening form (Appendix SB in supporting information).

The screening of titles and abstracts was carried out independently by two review authors (M.K. and O.B.) after completing calibration exercises ($n = 4119$). We compiled full texts for all titles and abstracts that met or were judged as potentially eligible to meet our inclusion criteria by at least one reviewer ($n = 309$); full texts were also screened independently by the same two reviewers. In total, 154 studies were included in the review (29 validation studies reviewed for the primary objective and 125 reviewed for the overview of cognitive assessment use); details of the study screening and selection are presented in the Results section. The kappa coefficient of inter-rater agreement for the full-text screening was found to be 0.93. Appendix SC in supporting information provides a list of all excluded articles at the full-text stage along with the reason for exclusion. Disagreements at each stage were resolved through discussions or by the corresponding author as needed.

2.4 | Data extraction

After completing calibration exercises, three teams of two reviewers extracted data in duplicate and independently using a pilot-tested form. They resolved disagreements through discussion or by referring to the corresponding author as needed.

We extracted the name of every cognitive assessment tool developed, validated, or used; the date of publication of the study; the country of the study; domains measured; study design; setting (clinical or epidemiological); inclusion and exclusion criteria; source of funding of the study; conflict of interest of study authors; sample size; and population characteristics (sex, age, and education).

For studies validating cognitive assessment tools, we abstracted information on the validity and reliability methods used, on tool performance, reported cut-offs, sensitivity and specificity, and whether normative data were provided. To provide a comprehensive mapping of validation efforts, we extracted all available information related to validity and reliability. We collected information on content, construct, and criterion validity. For reliability, we assessed if any reliability measure is reported: internal reliability (Cronbach alpha), test–retest reliability, inter-rater reliability, split-half reliability, or alternative forms of reliability. We further guided our data abstraction strategy by the Standards for Reporting Diagnostic accuracy studies[19] complemented with the following two best practice guidelines: *The International Guidelines for Translating and Adapting Tests* (Second Edition), published by the International Test Commission[20] and the *2014 Standards for Educational and Psychological Testing*[21] to collect information on translation, adaptation, and assembly given their importance for assessing applicability, reproducibility, and performance of cognitive tools. We collected information on translation procedures (forward translation only, backward translation only, forward and backward together, or if the translation was not applicable or unspecified) and on scale adaptation and development. We also abstracted any information on factors reported to influence tool performance. For studies assessing criterion validity (21 studies), we assessed methods and reporting quality using the Standards for Reporting of Diagnostic Accuracy Studies,[19] through evaluations and discussions by two reviewers (M.K. and M.E.). We noted some recurrent limitations (including small sample size, convenience sampling, and recruitment of controls from clinical settings); two studies reported sample size estimations,[22,23] and 12 established criterion validity against clearly described clinical diagnosis.[22–33] No limitation was deemed significant enough to warrant exclusion and we reviewed all identified studies, presenting their methods and variability in the Results and Discussion sections.

For the secondary objective of assessing studies using cognitive assessment tools, we abstracted data on whether the tool was self-developed or a previously developed validated instrument. Additionally, we evaluated whether the instruments' validity and language were reported.

3 | RESULTS

3.1 | Screening and study selection

The results of our search and selection process are presented in the PRISMA flowchart (Figure 1). The search on Medline, Embase, and APA PsycInfo yielded 5272 records. After removing duplicates, the number of articles retrieved for title and abstract screening was 4119. Of those, 3802 records did not meet the inclusion criteria, and 8 were not retrieved, leaving 309 records for full-text screening. A total of 161 records were excluded for not meeting the inclusion criteria (Figure 1). The reasons for exclusion are detailed in Appendix SC. Six additional articles were identified from the reference lists of the studies included. Thus, 154 studies were retained for analysis and review. Of those, 29 studies focused on developing or testing the validity of a tool, and 125 involved using a cognitive tool.

3.2 | Validated Arabic cognitive assessment tools

Across the 29 identified validation studies, validity properties were reported for 20 cognitive assessments (15 individual scales and five cognitive batteries). Below, we describe the general characteristics of validation studies, identified validated tools, validity methods, and tool performance, with a focus on tools validated in more than one study/setting.

3.2.1 | Characteristics of validation studies

The characteristics of the 29 studies examining the validity of a cognitive assessment tool are presented in Table 1. These validation studies were published between 2008 and 2023, with 12 (41.4%) published in

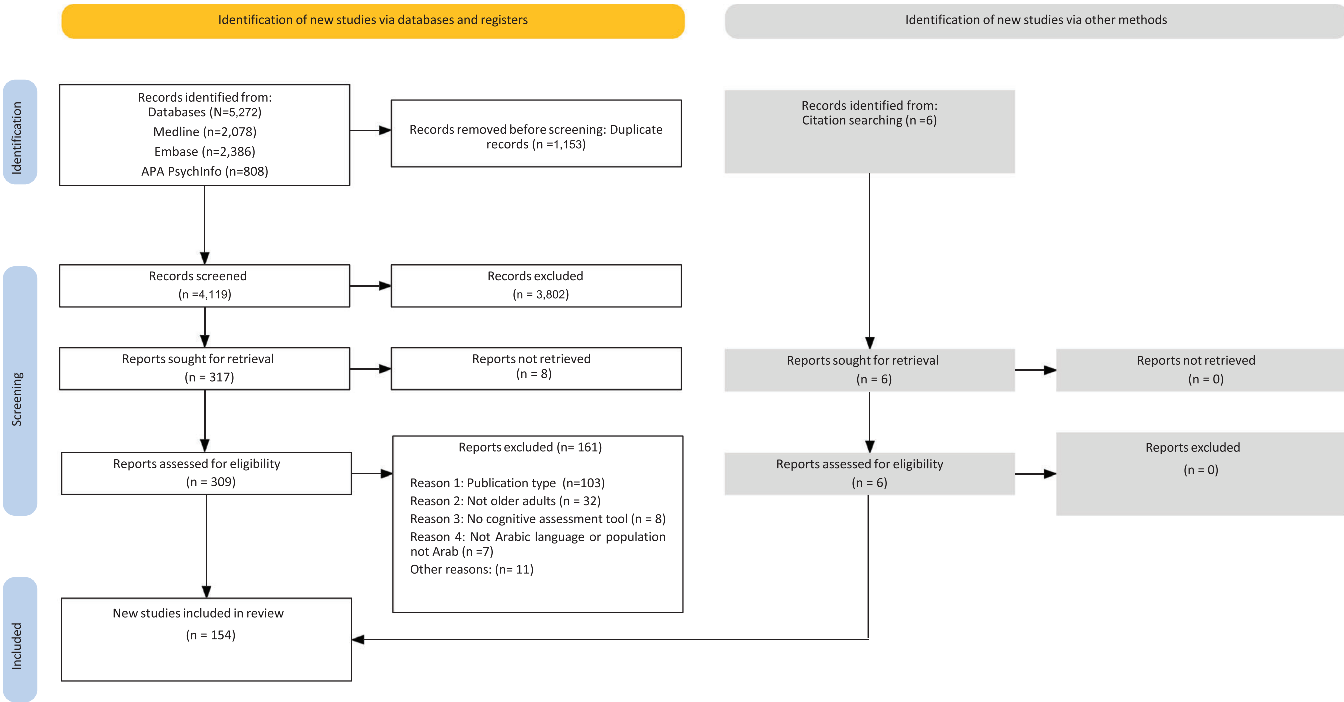

FIGURE 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of selected studies for inclusion.

the past 5 years or after 2018. Validation studies among Arab-speaking population were predominantly from Lebanon ($n = 9$) and Egypt ($n = 9$), with others from Tunisia ($n = 4$), Morocco ($n = 2$), Qatar ($n = 1$), Oman ($n = 1$), Jordan ($n = 1$), Israel ($n = 1$), and the United States ($n = 1$). The sample size ranged between 78 and 1010; details about the samples' age, sex, and education composition are presented in Table 1. Regarding the study setting, seven were conducted in a clinical setting, eleven were conducted in mixed settings (clinical settings with geriatric clubs, organizations, or nursing homes), two were conducted exclusively in geriatric/social centers, six studies were conducted in a community setting, and one in a mixed clinical/community setting. In two studies, the setting was not explicitly mentioned.

3.2.2 | Validated tools

Validity properties (content, criterion, or construct validity) were established for 15 individual cognitive assessment tools and five cognitive batteries (details are included in Table 2 and a summary is presented in Figure 2). Nine tools involved global cognitive screeners: the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), Rowland Universal Dementia Assessment Scale (RUDAS), Addenbrooke's Cognitive Examination III (ACE III), the shorter ACE III version (the mini-ACE), Cognitive Subscale of the Alzheimer's Disease Assessment Scale (ADAS-Cog), Dementia Arabic Scale (DAS), the Mini-cog, and the Brookdale Cognitive Screening Test (BCST). There were only two validated domain-specific tools (episodic memory): the Test of Nine Images 93 (A-TNI93) and the 5 Words Test (5WT). The five validated cognitive batteries were the 10/66 Dementia Research Group (DRG) diagnostic assessment for dementia (10/66), Consortium to Establish a Registry for Alzheimer's Disease neuropsychological battery – Arabic version (CERAD-ArNB), Dementia Screening Battery-100 (DSB-100), DemeGraph, and Ain Shams Cognitive Assessment (ASCA) scale. In addition, four tools—the Arabic Informant Questionnaire on Cognitive Decline in the Elderly (A-IQCODE), Clinical Dementia Rating (CDR) scale, Alzheimer's Questionnaire (AQ), and the Ascertain Dementia 8-item Informant Questionnaire (AD8)—were informant based. We note that the 10/66 and DemeGraph batteries include both informant and participant report scales.

3.2.3 | Validity methods

Translation, adaptation, and development

All 29 studies explicitly reported that the cognitive tests were administered in Arabic (Table 2). Ten studies used formal Arabic, while 11 used local dialects and eight did not specify the form of Arabic used (formal or dialect). Overall, 17 studies performed both tool adaptation and translation, six studies only performed translation, and three only performed adaptation. Regarding translation, forward and backward procedures were the most used ($n = 16$),[22,24–28,31–40] followed by forward translation alone ($n = 2$).[23,29] For five validation studies, translation procedures were not clearly reported.[30,41–44] Detailed descriptions on tool adaptation were reported in 12 studies.[23,24,26,30,32,34–36,43–46] In addition, three studies developed a new tool. The DSB-100, a cognitive battery with 10 subsets, was developed based on adapting several existing cognitive tests to the Tunisian context (including the token test, naming test, and forward and backward names span, a

TABLE 1 General characteristics of the validation studies.

Article	Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
Cognitive Assessment Tools										
SS[1] Hayek et al.,2020	UK**	MOCA	Cross-Sectional	164	Community	Lebanon	Mean ± SD: 70.10 ± 6.9 60–69 (50.62%) 70–79 (38.41%) 80+ (10.97%)	Males: 41.5% Females: 58.5% Ratio (M/F): 6/4	Primary (25.6%) Complementary (23.78%) Secondary (23.17%) University (27.53%)	Education Linguistic problems Age (limited effect) Gender (limited effect)
Abdel Rahman & El Gaafary,2009	UK*	MOCA	Stage I: Validation Stage II: Cross-Sectional	Stage I: 184 Stage II: 268	Geriatric Club	Egypt	Stage 1: Mean ± SD (years): 64.5 ± 6.8 Range: 60-83 Stage 2: Mean ± SD: 66.8 ± 5.05 Range: 60-76	Stage 1: Females: 90 (49%) Males: 94 (51%) Stage 2: Males: 146 (54.5%)	Stage 1: Illiterate: N = 1 High School Education: N = 90 Stage 2: High level of Education (secondary school to university): Majority	Not reported
Khatib et al.,2022	NF*	MOCA	Validation	106	Clinical Setting (controls were family members who accompanied the patient)	Morocco	<65 = 58 (54.7%) ≥65 = 48 (45.3%)	Males: 53 (47.3%) Females: 59 (52.7%)	Primary: 50 (44.6%) Secondary: 36 (32.1%) University: 26 (23.2%)	Education Age
Adel Saleh et al.,2019	NF**	MOCA-B	Case-Control	205	Geriatric Social Centre	Egypt	Mean ± SD Control Group: 65.9± 5.02 Major Neurocognitive Disorder: 72.11 ±7.64 Mild Neurocognitive Disorder: 67.72 ± 6.31	Ratio (M/F) Control group: 26/86 Major neurocognitive disorder: 27/27 Mild neurocognitive disorder: 17/22	Mean ± SD (years) Control group: 12.80 ± 4.38 Major neurocognitive disorder: 4.24 ± 5.84 Mild neurocognitive disorder: 9.74 ± 5.48	Low education
Wrobel & Farrag,2008	F*	MMSE	Validation	200	Community	USA (Arab Americans)	Mean ± SD: 69.06 ± 6.4 Range: 60-92	Males: N = 95 Females: N = 105	Range: 0 – 20 (years) Little or no formal education: Majority (69% reporting 3 years or less) Median: 2	Education Ethnicity (mentioned, not investigated)

(Continues)

TABLE 1 (Continued)

Article	Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
Rami et al., 2022	NF**	MMSE	Validation	80	Clinical (Outpatient Neurology Department)	Morocco	Dementia patients: 60–69: 8 (36.4%) 70–79: 5 (22.7%) ≥80: 9 (40.9%) Healthy controls: 60–69: 31 (53.4%) 70–79: 18 (31.0%) ≥80: 9 (15.5%)	Dementia patients: Males: 17 (77.3%) Females: 5 (22.7%) Healthy controls: Males: 35 (60.3%) Female: 23 (39.7%)	Dementia patients: Illiterate: 12 (54.5%) Illiterate non-formal Education: 3 (13.6%) Primary: 4 (18.2%) Middle school: 1 (4.5%) High school: 2 (9.1%) Healthy controls: Illiterate: 31 (53.4%) Illiterate non-formal Education: 8 (13.8%) Primary:10 (17.2%) Middle School: 3 (5.2%) High School: 5(8.6%) University: 1 (1.7%)	Education Cultural factors (mentioned, not investigated) Age (mentioned, not investigated) Social class (mentioned not investigated)
SS[2] Elkholy et al, 2018	UK**	MMSE	Cross-Sectional	159	Clinical and Elderly Institutions (Relatives of Patients in the Geriatric and Ophthalmology Wards, Patients Attending Geriatrics Outpatient Clinic and Geriatric Clubs)	Egypt	60+	Not specified	Illiterate and low education (≤ 9 years education): N = 58 High education (> 9 years education): N = 101	Education

(Continues)

TABLE 1 (Continued)

Article	Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
Dahbour et al., 2021	NF**	MMSE	Cross-sectional	250	Clinical (outpatient clinics) and Community	Jordan	Mean ± SD: 63 ± 8.4 Range: 50–86	Males: 111 (47%) Females: 125 (53%)	Illiteracy: 8.1% Up to high school level: 22% College (Beyond high school level): 69.9%	Education Age
Bellaj et al., 2008	UK*	A-MMSE	Validation	186	Unclear	Tunisia	Mean ± SD Normal: 66.11 ± 9.53 MCI: 64.93 ± 9.56 Dementia: 65.54 ± 8.96	Males: Normal: 41% MCI: 47% Dementia: 57% Females: Normal: 59% MCI: 53% Dementia: 43%	Mean ± SD (years): Normal: 6.34 ± 5.92 MCI: 7.46 ± 6.68 Dementia: 6.38 ± 6.86	Education Age
El-Hayeck et al., 2019	F**	A-MMSE (GTD-USJ)	Cross-Sectional	1,010	Community	Lebanon	55–59: 17.8% 60–64: 15.5% 65–69: 19.6% 70–74: 14.2% 75 +: 32.9%	Males: 50.9 % Females: 49.1%	< 6 years of education: 33.7% 6–12 years of education: 35.8% > 12 years of education: 30.5%	Age (mentioned, not investigated) Education (mentioned, not investigated) Ethnicity (mentioned, not investigated) Gender (mentioned, not investigated)
Inzelberg et al., 2007	F**	MMSE BCST	Cross-sectional	266	Community	Israel	Mean ± SD: 72.4 ± 5.5	Males: 158 (59.4%) Females: 108 (40.6%)	Mean ± SD (years): 4.0 ± 3.6 Range: 0–20 The distribution of education levels for each gender was as follows: Education level 1: Males: 72 (46%) / Females: 90 (84%) 2: Males: 62 (40%) / Females: 17 (15%) 3: Males: 18 (12%) / Females: 1 (0.1%)	Age Education Gender (for both tests)

(Continues)

TABLE 1 (Continued)

Article	Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
Albanna et al., 2017	F**	MMSE-2 Mini-Cog	Validation	134	Clinical: Primary Health Care Centers and Outpatient Services	Qatar (including non-Qataris)	Mean ± SD: 74.8 6± 7.71 60–64: 15 (11.19%) 65–69: 18 (13.42%) 70–74: 33 (24.63%) 75–79: 28 (20.9%) 80−-84: 22 (16.42%) 85–89: 10 (7.46%) 90+: 4 (2.99%) Unknown: 4 (2.99%)	Males: 83 (61.9%) Females: 51 (38.1%)	No school: 83 (61.9%) Intermediate: 13 (9.7%) Secondary: 10 (7.46%) College and above: 17 (12.69%) Unknown: 11 (8.21%)	Age Education Literacy Inability to read, write, draw Cultural factors (subcultural effects, mentioned, not investigated) Social class (mentioned, not investigated) Language (mentioned, not investigated) Gender (mentioned, not investigated)
		MMSE-2 & Mini-Cog		113						Inability to read, write, draw Cultural factors (subcultural effects)
SS[1] Abou-Mrad et al.,2017	F**	AD-8 MOCA MMSE 3MS BVMT-R LDS CLNT Fluency	Cross-Sectional	164	Community	Lebanon	Mean ± SD: 70.10 ± 6.91 Range: 60–87 60–64: 42 (25.6%) 65–69: 41 (25%) 70–74: 32 (19.51%) 75–79: 31 (18.9%) 80+: 18 (10.97%)	Males: 68 (41.5%) Females: 96 (58.5%)	Primaire (5–8): 42 (25.6%) Complémentaire (9–11): 39 (23.78%) Secondaire (12 and above without college degree): 38 (23.17%) Universitaire (BA/BS degree and above): 45 (27.43%)	Heightened test anxiety or reticence to participate (mentioned, not investigated) Cultural factors (mentioned, not investigated) Education Age Gender

(Continues)

TABLE 1 (Continued)

Article		Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
SS[3]	Qassem et al., 2020	F**	ACE III	Validation	78	Clinical and Elderly Institutions (Patients: Private and Public Memory Clinics, Neurology and Psychiatry Outpatient Departments Controls: Patients' Relatives, Friends, or the General Public through Cultural Centers and Elderly Clubs)	Egypt	Mean ± SD MCI: 74.83 ± 8.59 Controls: 72.48 ± 8.57	MCI: Males: N = 10 Females: N = 14 Controls: Males: N = 30 Female: N = 24	MCI: Elementary: N = 2 Preparatory: N = 1 Secondary: N = 4 Diploma: N = 2 University degree: N = 15 Controls: Elementary: N = 8 Preparatory: N = 5 Secondary: N = 10 Diploma: N = 6 University degree: N = 25	Education (mentioned, not investigated)
	Qassem et al., 2020 (2)	F**	ACE III	Validation	80	Clinical and Elderly Institutions (Patients: Private and public Memory Clinics, Neurology and Psychiatry Outpatient Departments Controls: Patients' relatives, friends, or the General Public through Cultural Centers and Elderly Clubs)	Egypt	Mean ± SD Dementia: 77.05±7.03 Controls: 74.67±8.22	Dementia: Males: N = 17 Females: N = 20 Controls: Males: N = 21 Females: N = 22	MCI: Elementary: N = 3 Preparatory: N = 2 Secondary: N = 11 Diploma: N = 2 University degree: N = 19 Controls: Elementary: N = 4 Preparatory: N = 4 Secondary: N = 10 Diploma: N = 4 University Degree: N = 21	Education (mentioned, not investigated)

(Continues)

TABLE 1 (Continued)

Article	Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
Qassem et al.,2021	F**	Mini-ACE	Validation	80	Clinical and Elderly Institutions: (Patients: Private and Public Memory Clinics, Neurology and Psychiatry Outpatient Departments Controls: Patients' Relatives, Friends, or the General Public through Cultural Centers and Elderly Clubs)	Egypt	Mean ± SD Dementia: 77.05 ± 7.03 Controls: 74.67 ± 8.22	Dementia: Males: N = 17 Females: N = 20 Controls: Males: N = 21 Females: N = 22	MCI: Elementary: N = 3 Preparatory: N = 2 Secondary: N = 11 Diploma: N = 2 University Degree: N = 19 Controls: Elementary: N = 4 Preparatory: N = 4 Secondary: N = 10 Diploma: N = 4 University degree: N = 21	Type of sample (mentioned, and not investigated) Education (mentioned, and not investigated)
Qassem et al., 2015	NF**	ACE III	Validation	139	Clinical and Elderly Institutions (Cultural centers and elderly clubs, relatives, and friends of patients in Outpatient Clinic at the Neurology and Gerontology Departments).	Egypt	Median (IQR):71 (13.75) Range: 60-93	Males: 57.1% Females: 42.9%	Basic: 25% Secondary: 28.6% University: 46.4%	Age influenced only the performance on the category fluency
El-Hayeck et al., 2023	F**	A-TNI93 (GTD-USJ)	Cross-Sectional	332	Community	Lebanon	55-59: N = 39 60-64: N = 53 65-69: N = 53 70-74: N = 64 75+: N = 123	Males: N = 155 Females: N = 177	All participants are illiterate	Age Gender
Ben Jemaa et al.,2017	UK**	ADAS-Cog	Validation	182	Clinical (Neurology Department)	Tunisia	Mean ± SD AD: 69.40 ± 8.59 (min: 50, max: 89) N-AD: 69.06 ± 6.27 (min: 50, max: 87) Control: 68.15 ± 7.13 (min: 50, max: 95)	Males: AD: 13 N-AD: 16 Control: 58 Females: AD: 12 N-AD: 17 Control: 65	Mean ± SD (years): AD: 6.60 ± 6.65 N-AD: 6.11 ± 3.82 Control: 5.44 ± 5.74	Age Education Cultural factors (mentioned, not investigated)

(Continues)

TABLE 1 (Continued)

Article		Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
SS[4]	Chaaya et al., 2016	F**	A-RUDAS	Validation	232	Clinical and Elderly Institutions (Social Organizations for the Elderly, Hospital-Based Neurological and Geriatric Clinics, and Community-Based Primary Care Clinics)	Lebanon	Mean ± SD: 79.1 ± 8.1 65-74: N = 74 75-84: N = 96 85+: N = 62	Males: 1/3 Females: 2/3	No formal education: 51.7% Primary but not secondary education: 29.7% Completed secondary: 11.64% University education: 6.9%	Education
	Phung et al.,2015	F**	A-IQCODE 16	Validation	236	Clinical and Elderly Institutions (Social Organizations for the Elderly, Hospital-Based Neurological and Geriatric Clinics, and Community-Based Primary Care Clinics)	Lebanon	Mean ± SD Dementia: 81.8±7.7 Normal cognition: 77.0±7.9	Dementia: Males: 29 (31.2%) Females: 64 (68.8%) Normal cognition: Males: 54 (37.8%) Female 89 (62.2%)	Dementia: No formal education (illiterate, read and write): 56 (60.2%) Formal education (primary education and above): 37 (39.8%) Normal cognition: No formal education (illiterate, read and write): 62 (43.4%) Formal education (primary education and above): 81 (56.6%)	Education (investigated but no effect)
Mrabet Khiari et al.,2008		UK*	5WT	Validation	134	Unclear	Tunisia	Mean ± SD Normal: 66.55 ± 8.65 Non DTA: 67.39 ± 8.82 DTA: 71.37 ± 8.21	Males: Normal: N = 52 Non DTA: N = 15 DTA: N = 15 Females: Normal: N = 24 Non DTA: N = 16 DTA: N = 12	Mean ± SD (years): Normal: 6.88 ± 6.47 Non DTA: 4.90 ± 7.31 DTA: 4.81 ± 6,29	Age Education

(Continues)

TABLE 1 (Continued)

Article	Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
Farghaly et al.,2021	UK**	DAS	Validation	240	Clinical (university hospital outpatient clinic or inpatient ward)	Egypt	Mean ± SD Dementia: 68.45 ± 8.5 Controls: 66.43 ± 12.2 Median Dementia: 68 Controls: 63.5 *Range* Dementia: 50-87 Controls: 50-89	Dementia: Males: N = 73 Females: N = 47 Control: Males: N = 69 Females: N = 51	Dementia: Illiterate: 78 (65%) Basic/secondary: 29 (24.2%) Higher education: 13 (10.8%) Controls: Illiterate: 78 (65%) Basic/secondary: 23 (19.2%) Higher education: 19 (15.8%)	Not reported
Karam et al., 2018	F*	AD8 AQ CDR	Validation	127	Clinical and elderly institutions (outpatient clinics, inpatient units, nursing home residents)	Lebanon	Mean ± SD 81.9 ± 7.8	Females: 81(61.4%) Males: 51 (38.6%)"	Illiterate: 19 (14.4%) 6 years or less: 53 (40.1%) 7-10 years: 22 (16.7%) 11-13 years: 19 (17.4%) Started university: 1 (0.8%) University graduate: 18 (13.6%)	Not reported
Cognitive Assessment Batteries										
Bellaj et al., 2017	UK**	DSB-100	Validation	201	Clinical (Cases: Memory Clinics and Controls: Relatives of the Patient)	Tunisia	Mean ± SD Dementia patients: 69.38 ± 7.75 (min: 55, max: 89) Healthy controls: 67.80 ± 7.25 (min:55-max:90)	Males: Dementia patients: 15 (35.71%) Healthy control: 78 (49.06%) Females: Dementia patients: 27 (64.29%) Healthy control: 81 (50.94%)	Mean ± SD (yrs): Dementia patients: 5.93 ± 6.84 (min: 0, max: 21) Healthy controls: 5.47 ± 6.64 (min: 0, max: 21)	Age Education

(Continues)

TABLE 1 (Continued)

Article		Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
SS[4]	Nielsen et al., 2015	F**	Deme-Graph: RUDAS & IQCODE	Validation	225	Clinical and Elderly Institutions (Social Organizations for the Elderly, Hospital-Based Neurological and Geriatric Clinics, and Community-Based Primary Care Clinics)	Lebanon	Mean ± SD Dementia patients: 81.9±7.5 Controls: 77.0±7.5	Females: Dementia patients: 61 (67.8%) Controls: 83 (61.5%)	Dementia patients: No formal education: 55 (61.1%) Primary education: 16 (17.8%) Intermediate education: 11 (12.2%) Secondary education: 7 (7.8%) University and above: 1 (1.1%) Controls: No formal education: 61 (45.2%) Primary education: 23 (17.0%) Intermediate education: 19 (14.1%) Secondary education: 18 (13.3%) University and Above: 14 (10.4%)	IQCODE: affected by informant characteristics such as depression and anxiety, the level of a carer's burden, and the quality of the relationship between the informant and older participant (mentioned, not investigated)
	Phung et al.,2014	F*	10/66 Dementia Research Group (DRG) diagnostic assessment for dementia	Validation	244	Clinical and Elderly Institutions (Social Organizations for the Elderly, Hospital-Based Neurological and Geriatric Clinics, and Community-Based Primary Care Clinics)	Lebanon	65-74: 78 (31.9%) 75-84: 99 (40.5%) 85-100: 67 (27.5%)	Females: 158 (64.8%) Males: 86 (35.2%)	No formal education: 123 (50.4%) Primary: 45 (18.4%) Intermediate: 32 (13.1%) Secondary: 27 (11.1%) University & above: 17 (7.0%)	GMS: Education and depression (mentioned, not investigated)

(Continues)

TABLE 1 (Continued)

Article	Funding & Conflict of Interest	Tool	Study Design	Sample Size	Setting	Country	Age	Gender	Education	Factors reported to influence performance
SS[2] Elkholy et al.,2020	NF**	ASCA	Cross-Sectional	159	Clinical and Elderly Institutions (Relatives of Patients in the Geriatric and Ophthalmology Wards, Patients Attending Geriatrics Outpatient Clinics, and Geriatric Clubs)	Egypt	60+	Males: N = 78 Females: N = 92	Illiterate and low education (≤ 9 years education): N = 58 High education (> 9 years education): N = 101	Education
Alobaidy et al.,2017	PF**	CERAD-ArNB	Cross-Sectional	150	Clinical: Tertiary Referral Center	Oman	50-64: 88 (70.4%) ≥ 65: 37 (29.6%) Range: 50-83	Males: 65 (52%) Females: 60 (48%)	0-5: 77 (61.6%) ≥ 6: 48 (38.4%) Range: 0-20	Education Age Gender

Abbreviations: 3MS, Modified Mini Mental State; 5WT, 5 Words Test; ACE III, Addenbrooke's Cognitive Examination III; AD, Alzheimer's disease; AD8, Alzheimer's Disease 8-item Questionnaire; ADAS-Cog, Cognitive Subscale of the Alzheimer's Disease Assessment Scale; A-IQCODE 16, Arabic Informant Questionnaire on Cognitive Decline in the Elderly; A-MMSE, Arabic Mini-Mental Status Examination; A-MMSE(GTD-USJ), Arabic Mini-Mental Status Examination – Groupe de Travail sur les Démences de l'Université Saint Joseph; AQ, The Alzheimer's Questionnaire; A-RUDAS, Arabic Rowland Universal Dementia Assessment Scale; ASCA, Ain Shams Cognitive Assessment Scale; A-TNI93 (GTD-USJ), Arabic Test des Neuf Images du 93-Groupe de Travail sur les Démences de l'Université Saint Joseph; BCST, Brookdale Cognitive Screening Test; BVMT-R, Brief Visuospatial Memory Test-Revised; CDR, Clinical Dementia Rating scale; CERAD-ArNB, Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Battery-Arabic version; CLNT, Cross-Linguistic Naming Test; DAS, Dementia Arabic Scale; DSB-100, Dementia Screening Battery-100; F, funded; LDS, Lebanese Digit Span; MCI, mild cognitive impairment; Mini-ACE, Mini Addenbrooke's Cognitive Examination; MMSE, Mini-Mental Status Examination; MMSE-2, Mini-Mental State Examination Second Edition; MoCA, Montreal Cognitive Assessment; MoCA-B, Montreal Cognitive Assessment-Basic; N-AD, non-Alzheimer's dementia; NF, not funded; PF, partially funded; SD, standard deviation; SS, same sample; UK, unknown.

*Conflict of interest not declared.

**No conflict of interest.

TABLE 2 Psychometric properties of validated cognitive tests.

Article	Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct (EFA & CFA), Criterion)	Validation Psychometrics: PPV	Validation Psychometrics: NPV	Validation Psychometrics: SEN/SPEC	Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
Cognitive Assessment Tools											
SS[1] Hayek et al., 2020	MOCA	Executive Functions and Visuospatial Abilities, Denomination, Memory, Attention, Concentration and Working Memory, Language, Abstraction, Delayed Recall, Orientation	164 Only 24 followed up for Test-retest Lebanon	Formal Arabic	ADAP	Yes	Construct validity (external nomological validity) *				Cronbach alpha: 0.82 Test–retest: 0.70 Inter-rater (Kappa): 0.61
Abdel Rahman & El Gaafary,2009	MOCA	Visuospatial, Executive Function, Attention, Reading Digits, Serial 7 Subtraction, Language, Recent Memory, Delayed Recall, Orientation	Stage I: 184 Stage II: 268 Egypt	Arabic, no type reported	TADAP	No	Content validity Criterion validity (against CAMCOG)			Cut-off: 26 (MCI screening) Sensitivity: 92.30% Specificity: 85.70%	Cronbach alpha: 0.83 Test–retest reliability: 0.92
Khatib et al.,2022	MOCA	Visuospatial/Executive Functions, Naming, Memory, Attention, Language, Abstraction, Recall, and Orientation.	106 Morocco	Darija, Tamazight in its three variants (Tachelhit, Tarifit, Atlas Tamazight), and Arabic	TADAP	No	Criterion validity (against clinical diagnosis) Construct validity (external nomological validity) *	92.10%	90.90%	Cut-off: 25 (dementia screening) Sensitivity: 88.50% Specificity: 93.80%	Cronbach alpha: 0.87
Adel Saleh et al.,2019	MOCA-B	Visual Perception (Superimposed Objects), Executive Functioning (Simplified Alternating Trail Making), Word Similarity (Problem Solving), Language (Fruit Fluency, Animal Naming), Attention (Modified Digit Stroop), Memory (Five-Word Delayed Recall), and Orientation (Time and Place)	205 Egypt	Formal Arabic	T	No	Content validity Criterion validity (against clinical diagnosis) Construct validity (known group validity) *	Mild NCD: 97.7% Major NCD: 92.4%	Mild NCD: 94.0% Major NCD: 96.7%	Dementia: Cut-off: 21/22 (all subjects) Sensitivity: Mild NCD: 92.5% Specificity: Mild NCD: 98.2% Cut-off: 16/17 (all subjects) Sensitivity: Major NCD: 90.7% Specificity: Major NCD: 97.4% cutoff: 18/19 (low education subjects) Sensitivity: Mild NCD: 87% Specificity: Mild NCD: 100% Cut-off: 13/14 (low education subjects) Sensitivity: Major NCD: 92.5% Specificity: Major NCD: 100%	Cronbach alpha: 0.915

(Continues)

TABLE 2 (Continued)

Article	Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct Validity, Construct (EFA & CFA), Criterion)	Validation Psychometrics: PPV	Validation Psychometrics: NPV	Validation Psychometrics: SEN/SPEC	Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
Wrobel & Farrag,2008	MMSE	Orientation, Immediate Registration, Attention and Calculation, Short-Term Recall, Language, and Visual Construction	200 USA (Arab Americans)	Arabic Version, no type reported	TADAP	No	Criterion validity (against IQCODE) Construct validity (known group validity and external nomological validity)*	67% (cut-off: 23) 69.5% (cut-off: 22, for low education) 60.7% (cut-off: 23 high education)		Cut-off: 23 (cognitive decline screening) Sensitivity: 73% Specificity: 60% Cut-off toff: 22 (low education) Sensitivity: 79% Specificity: 27% Cut-off: 23 (high education) Sensitivity: 52% Specificity: 82%	Cronbach alpha: 0.89
Rami et al., 2022	MMSE	Orientation, Registration, Attention, Recall, Naming, Repeating, Following Commands, Following Picture, Sentence Construction, Figure Copying	80 Morocco	Moroccan-Arabic	TADAP	No	Criterion validity (against clinical diagnosis) Construct validity (Known Group validity)*			Dementia screening: Cut-off: 21/22 (literate elderly-dementia) Sensitivity: 85.7% Specificity: 94.7% Cut-off: 18/19 (illiterate elderly-dementia) Sensitivity: 92.9% Specificity: 95.0%	Cronbach alpha: literate: 0.764 Illiterate: 0.761 Test–retest reliability: 0.78
SS[2] Elkholy et al, 2018	MMSE	Orientation, Registration, Attention and Calculation, Recall, Language (Repetition and Complex Command), and Visuospatial Functions	159 Egypt	Not Specified	ADAP	No	Criterion validity (against DSM IV)	44.03%	95.70%	Cut-off: 23 (dementia diagnosis) Sensitivity: 95% Specificity: 73.83% Cut-off: ≤ 21 (low education): Sensitivity: 91% Specificity: 33.33% Cut-off: ≤ 22 (high education) Sensitivity: 94% Specificity: 70.2%	
Dahbour et al., 2021	MMSE	Orientation, Registration, Attention, Language, and Recall	250, Jordan	Arabic version, no type reported	T (not clear how)	Yes	Construct validity (external nomological validity) *				

(Continues)

TABLE 2 (Continued)

Article	Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct Validity (EFA & CFA), Criterion)	Validation Psychometrics			Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
								PPV	NPV	SEN/SPEC	
Bellaj et al., 2008	A-MMSE	Orientation, Recording Information, Attention and Mental Arithmetic, Word Recall, Language and Constructive Praxis	186 Tunisia	Tunisian-Arabic	TADAP	Yes	Criterion validity (against clinical diagnosis) Construct validity (external nomological validity) *			Cut-off: 26 (dementia screening) Sensitivity: 95% Specificity: 82% Cut-off: 24 (dementia diagnosis): Sensitivity: 71% Specificity: 99%	Cronbach alpha: 0.72 Test–retest: 0.95
El-Hayeck et al., 2019	A-MMSE(GTD-USJ)	Orientation, Registration, Attention, Recall, Language, and Constructional Praxis	1,010 Lebanon	Formal Arabic	TADAP	No	Criterion validity (against CDR based clinical diagnosis, performed for suspected cases based on lower MMSE scores*) Construct validity (external nomological validity) *	PLR: 7.6	NLR: 4.5	Cut-off: 23/24 (dementia screening) Sensitivity: 80% Specificity: 89.44%	Cronbach alpha: Total score: 0.71 For orientation: 0.64 for registration: 0.65 For calculation-attention: 0.8 For recall: 0.52 For language: 0.37 Test–retest: 0.72 Inter-rater: 0.89
Inzelberg et al., 2007	MMSE BCST	Orientation, Language, Memory, Attention, Naming, Abstraction, Concept Formation, Attention, Praxis, Calculation, Right Left Orientation, and visuospatial Orientation	266 Israel	Arabic version, no type reported	T (not clear how)	Yes	Construct validity (external nomological validity) *				
Albanna et al., 2017	MMSE-2	Orientation Recall Attention Calculation Language Processing Constructional Praxis	134 & 113 Qatar (including non-Qataris)	Formal Arabic	TADAP	No	Criterion validity (against clinical diagnosis based on DSM-IV-TR) Construct validity (external nomological validity) *			Dementia screening: cut-off: 18/19 (MMSE adjusted scores) Sensitivity: 60.9% Specificity: 59.5% Cut-off: 21/22 (MMSE *t* scores) Sensitivity: 59.1% Specificity: 68.7%	
	Mini-Cog	Visuo-Constructive Abilities Including Praxis and Executive Higher Functions								Cut-off: 1.5 Sensitivity: 92.9% Specificity: 34.3%	
	MMSE-2 & Mini-Cog									Cut-off: 20/21 Sensitivity: 71.4% Specificity: 61.6%	

(Continues)

TABLE 2 (Continued)

Article		Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct (EFA & CFA), Criterion)	Validation Psychometrics PPV	NPV	SEN/SPEC	Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
SS[1]	Abou-Mrad et al., 2017	AD-8	Memory and Other Thinking Abilities	164 Only 24 followed up for test-retest Lebanon	Literary Arabic	TADAP	Yes	None however, Construct validity (EFA) for all the tools together.				Cronbach alpha: 0.34 Test–retest: 0.74
		MOCA	Memory, Attention, Language, Construction, and Executive Functioning									Cronbach alpha: 0.82 Test–retest: Total (raw): 0.70 Total (education adjusted): 0.67
		MMSE	Attention, Immediate and Short-Term Memory, Orientation, Language, and Visuospatial Function									Cronbach alpha: 0.85 Test–retest: 0.67
		3MS	Tapping Memory, Attention, Language, Construction, and Executive Functioning									Cronbach's Alpha: 0.79 Test–retest: 0.70
		BVMT-R	Recall									Cronbach alpha: 0.85 Test–retest: 0.66/0.84/0.83
		LDS	Forward and Backward Auditory Attention									Cronbach alpha: 0.93 Test–retest: 0.75
		CLNT	Correct Naming, Phonemic Paraphasias, Semantic Paraphasias, And Circumlocutions of Images Grouped Into 5 Categories									Cronbach alpha: 0.75 Test–retest: 0.90
		Fluency	Verbal Fluency (Phonemic: Letter of The Alphabet/Semantic: Category)									Cronbach alpha: 0.48 Test–retest Phonemic fluency: 0.55 Semantic fluency: 0.57

(Continues)

TABLE 2 (Continued)

Article	Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct Validity, Construct (EFA & CFA), Criterion)	Validation Psychometrics			Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
								PPV	NPV	SEN/SPEC	
SS[3] Qassem et al., 2020	ACE III	Attention, Memory, Fluency, Language, Visuospatial ability	78 Egypt	Egyptian-Arabic	TADAP	No	Criterion validity (against clinical diagnosis*) Construct validity (known group validity)			Cut-off: 81 (MCI screening) Sensitivity: 75% Specificity: 82%	
Qassem et al., 2020 (2)	ACE III	Attention, Memory, Fluency, Language, and Visuospatial Processing	80 Egypt	Egyptian-Arabic	TADAP	No	Criterion validity (against clinical diagnosis*) Construct validity (known group validity)			Cut-off: 72 (dementia diagnosis) Sensitivity: 89% Specificity: 95%	
Qassem et al., 2021	Mini-ACE	Attention, Memory Registration, Fluency, Memory Recall, and Visuospatial Abilities	80 Egypt	Egyptian-Arabic	TADAP	No	Criterion validity (against clinical diagnosis) Construct validity (known group validity, and external nomological validity)			Cut-off: 18 (dementia screening) Sensitivity: 92% Specificity: 95%	
Qassem et al., 2015	ACE III	Attention and Orientation, Memory, Verbal Fluency, Language and Visuospatial Abilities	139 Egypt	Egyptian-Arabic	TADAP	Yes					
El-Hayeck et al., 2023	A-TNI93 (GTD-USJ)	Episodic memory	332 Lebanon	Not Mentioned	ADAP	Yes	Criterion validity (against CDR based clinical diagnosis, performed for suspected cases based on index test scores*) Construct validity (external nomological validity)*	41%	96.50%	Cut-off: FR $\leq$ 6 (dementia acreening) Sensitivity: 66.7% Specificity: 90.5% Cut-off: FR $\leq$ 6 & total recall $\leq$ 8 (dementia screening) Sensitivity: 70.8% Specificity: 88%	Test–retest: 0.63 Inter-rater: 0.29
Ben Jemaa et al.,2017	ADAS-Cog	Spoken Language Ability Comprehension of Spoken Language Word-finding Word Recall Naming of Objects and Fingers Orientation Commands Ideational Praxis Construction Praxis Word Recognition Recall of test instructions	182 Tunisia	Arabic Version, No Type Reported	TADAP	Yes	Criterion validity (against clinical diagnosis) Construct Validity (PCA, Known group validity*, and external nomological validity*)			Cut-off: 10 (AD diagnosis) Sensitivity: 84% Specificity: 91%	Cronbach alpha: 0.81 Test–retest: AD (r = 0.97) N-AD patients (r = 0.93) NC subjects (r = 0.87) Overall (r = 0.97)

(Continues)

TABLE 2 (Continued)

Article		Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct (EFA & CFA), Criterion)	Validation Psychometrics			Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
									PPV	NPV	SEN/SPEC	
SS[4]	Chaaya et al., 2016	A-RUDAS	Testing Memory (registration and recall) Body Orientation Praxis Drawing Judgment Language	232 Lebanon	Classical Arabic	T	No	Criterion validity (against clinical diagnosis based on DSM-IV) Construct validity (known group validity and external nomological validity)*	75 (cut-off: ≤ 23) 79 (cut-off: ≤ 22)		Cut-off: ≤ 23 (dementia screening) Sensitivity: 92% Specificity: 78% Cut-off: ≤ 22 Sensitivity: 83% Specificity: 85%	Cronbach alpha: 0.87
	Phung et al.,2015	A-IQCODE 16	NA	236 Lebanon	Arabic Version, no type reported	T	No	Content validity Criterion validity (against a clinician's dementia diagnosis) Construct validity (PCA, Known group validity*, and external nomological validity*)	91.5% (total) 93.1% (no formal education) 88.9% (formal education)	95.1% (Total)	Cut-off: > 3.34 (dementia screening) Sensitivity: 92.5% Specificity: 94.4% Cut-off: > 3.34 (no formal education) Sensitivity: 96.4% Specificity: 93.5% Cut-off: > 3.34 (formal education) Sensitivity: 86.5% Specificity: 95.1%	Cronbach alpha: 0.97
Mrabet Khiari et al.,2008		5WT	Episodic Memory	134 Tunisia	Tunisian-Arabic	TADAP	No	Criterion validity (against clinical diagnosis)* Construct validity (external nomological validity) *			Cut-off: STP ≤ 17 (AD diagnosis) Sensitivity: 81% Specificity: 86% Cut-off: ST ≤ 9 (AD diagnosis) Sensitivity: 70% Specificity: 93% Cut-off: SM ≤ 4 (AD diagnosis) Sensitivity: 70% Specificity: 98%	

(Continues)

TABLE 2 (Continued)

								Validation Psychometrics			
Article	Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct (EFA & CFA), Criterion)	PPV	NPV	SEN/SPEC	Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
Farghaly et al., 2021	DAS	Orientation to Time, Persons and Place, Memory Including Registration and Repetition, Attention, Executive Function, Speech and Language Including Category Fluency, Assessment of Ability of Reading, Comprehension and Writing, Judgment, and Social Cognition.	240 Egypt	Arabic Version, no type reported	D	No	Face & Content Validity Criterion validity (against MMSE, CASI, and clinical diagnosis*) Construct validity (contrasted group validity)	For dementia: Literate: 86% Illiterate: 95.5% For mild dementia: Literate: 94.5% Illiterate: 100% For moderate dementia: Literate: 90.5% Illiterate: 92% For severe dementia: Literate: 100% Illiterate: 96%	For dementia: Literate: 100% Illiterate: 88% For mild dementia: Literate: 100% Illiterate: 84% For moderate dementia: Literate: 96.5% Illiterate: 93% For severe dementia: Literate: 95.5% Illiterate: 94%	Cut-off: 95 (dementia diagnosis) Literate: Sensitivity: 100% specificity: 84% Cut-off: 68 (dementia) Illiterate: Sensitivity: 87% Specificity: 96% Cut-off: 95 (mild dementia) Literate: Sensitivity: 100% Specificity: 94% Cut-off: 68 (mild dementia) Illiterate: Sensitivity: 81% Specificity: 100% Cut-off: 62 (moderate dementia) Literate: Sensitivity: 97% Specificity: 90% Cut-off: 57 (moderate dementia) Illiterate: Sensitivity: 93% Specificity: 92% Cut-off: 35 (severe dementia) Literate: Sensitivity: 95% Specificity: 100% Cut-off: 35 (severe dementia) Illiterate: Sensitivity: 94% Specificity: 96%	Cronbach alpha: 0.888

(Continues)

TABLE 2 (Continued)

Article	Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct (EFA & CFA), Criterion)	Validation Psychometrics: PPV	NPV	SEN/SPEC	Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
Karam et al., 2018	AD8	NA	127 Lebanon	Arabic Version, no type reported	T	No	Criterion validity (against clinical diagnosis based on NINCDS-ADRDA criteria) Construct validity (external nomological validity) *	96.21%		Cut-off: 3 (discriminate cases from non-cases of dementia) Sensitivity: 96% Specificity: 97% Cut-off: 4 (discriminate cases from non-cases of dementia) Sensitivity: 95% Specificity: 98%	Cronbach alpha: 0.9
	AQ							100%		Cut-off: 8 (discriminate cases from non-cases of dementia) Sensitivity: 100% Specificity: 100%	Cronbach alpha: 0.94
	CDR							95%		Cut-off: ≥ 2 (A-CDR-SOB-discriminate cases from non-cases of dementia) Sensitivity: 97% Specificity: 93%	Cronbach alpha: 0.98
Cognitive Assessment Batteries											
Bellaj et al.,2017	DSB-100	Memory, Executive Functions, Praxis, Language, Attention, and Visuospatial Functions	201 Tunisia	Tunisian-Arabic	D	No	Construct validity (PCFA and external nomological validity *)				Cronbach alpha: 0.76 Test–retest: Patients ($r = 0.81$, $P < 0.001$) Controls ($r = 0.87$, $P < 0.001$) Inter-rater reliability: high

(Continues)

TABLE 2 (Continued)

Article		Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct (EFA & CFA), Criterion)	Validation Psychometrics: PPV	Validation Psychometrics: NPV	Validation Psychometrics: SEN/SPEC	Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
SS[4]	Nielsen et al., 2015	DemeGraph: RUDAS & IQCODE	RUDAS: Body Orientation, Praxis, Drawing, Judgment, Memory and Language IQCODE: The Short Form of The IQCODE Contains 16 Questions and Rates Changes in Cognitive Functioning of an Older Adult Over a Specified Time Frame	225 Lebanon	Classical Arabic	TADAP	No	Criterion validity (against clinical diagnosis based on DSM-IV-TR) Construct validity (known group validity)*	95%	91%	Dementia screening: IQCODE: cut-off: > 3.34/5 & RUDAS: < 23/30 Sensitivity: 86% Specificity: 97%	
	Phung et al.,2014	10/66 Dementia Research Group (DRG) diagnostic assessment for dementia	NA	244 Lebanon	Classical Arabic with Lebanese Modification in Language Parts	TADAP	No	Content validity Criterion validity (against clinician dementia diagnosis)	10/66: 92.9% CSI-D: 86.8% CERAD 10: 65.9% GMS: 74%		(Dementia diagnosis) 10/66 Cut-off $\geq$ 0.25592 Sensitivity: 92% Specificity: 95.1% CSI-D DFSCORE: Sensitivity: 92% Specificity: 90.3% CERAD 10 (cut-off: 3) Sensitivity: 91% Specificity: 67.4% GMS Sensitivity: 94% Specificity: 77.1%	
SS[2]	Elkholy et al., 2020	ASCA	Verbal Learning Short and Long-term Memory Cued Memory Short-term and Working Memory Visuospatial Functions Executive Functions Verbal Fluency Confrontation Naming	159 Egypt	Egyptian-Arabic	D	No	Content validity* Construct validity (Criterion group validity and external nomological validity*)				Test–retest reliability: adequate to high with $r \geq 0.7$

(Continues)

TABLE 2 (Continued)

Article	Tool	Domain	Sample Size & Country	Language	Method of development	Normative Data	Validity investigations (Content & Face Validity, Construct (EFA & CFA), Criterion)	Validation Psychometrics			Reliability tests (internal (Cronbach alpha), test–retest, inter-rater, split half, alternate forms)
								PPV	NPV	SEN/SPEC	
Alobaidy et al., 2017	CERAD-ArNB	MMSE: Memory, Concentration, Language, and Constructional Praxis CERAD Word List Learning, Recall, and Recognition Tests: Verbal Memory Verbal Fluency Test: Semantic Memory and Lexical Access Ability 15-Item Modified Boston Naming: Confrontational Naming Constructional Praxis Copy and Constructional Praxis Recall Tests: Constructional Ability and Visual Memory	150 Oman	Local Arabic	TADAP	Yes	Construct validity (external nomological validity)*				

Abbreviations: 5WT, 5 Words Test; 3MS, Modified Mini-Mental State; A, Arabic; ACE III, Addenbrooke's Cognitive Examination III; A-CDR-SOB, Arabic Clinical Dementia Rating Scale Sum of Boxes Score; AD, Alzheimer's disease; ADAP, adaptation; ADAS-Cog, Cognitive Subscale of the Alzheimer's Disease Assessment Scale; AD8, Alzheimer's Disease 8-item Questionnaire; A-IQCODE 16,: Arabic Informant Questionnaire on Cognitive Decline in the Elderly; A-MMSE (GTD-USJ), Arabic Mini-Mental Status Examination – Groupe de Travail sur les Démences de l'Université Saint Joseph; A-RUDAS, Arabic Rowland Universal Dementia Assessment Scale; AQ, The Alzheimer's Questionnaire; ASCA, Ain Shams Cognitive Assessment Scale; A-TNI93 (GTD-USJ), Arabic Test des Neuf Images du 93-Groupe de Travail sur les Démences de l'Université Saint Joseph; BCST, Brookdale Cognitive Screening Test; BVMT-R, Brief Visuospatial Memory Test-Revised; CAMCOG, Cambridge Cognition Examination; CASI, Cognitive Abilities Screening Instrument; CERAD-ArNB, Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Battery-Arabic version; CDR, Clinical Dementia Rating Scale; CDR-SB, Clinical Dementia Rating Scale-Sum of Boxes Score; CFA, Confirmatory Factor Analysis; CLNT, Cross-Linguistic Naming Test; CSI-D, Community Screening Instrument for Dementia; D, developed; DAS, Dementia Arabic Scale; DFSCORE, Discriminant Function Score; DSB-100, Dementia Screening Battery-100; DSM IV, Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition; DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition-Text Revision; EFA, exploratory factor analysis; FR, Free Recall; LDS, Lebanese Digit Span; MCI, Mild Cognitive Impairment; Mini-ACE, Mini Addenbrooke's Cognitive Examination; MMS, Mini-Mental State; MMSE, Mini-Mental Status Examination; MMSE-2, Mini-Mental State Examination Second Edition; MOCA, Montreal Cognitive Assessment; MOCA-B, Montreal Cognitive Assessment-Basic; GMS, Geriatric Mental State; N-AD, non-Alzheimer's disease dementia; NC, normal controls; NCD, neurocognitive disorder; NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association Criteria; NLR, negative likelihood ratio; NPV, negative predictive value; PCA, principal component analysis; PCFA, principal component factor analysis; PLR, positive likelihood ratio; PPV, positive predictive value; SEN, sensitivity; SPEC, specificity; T, translation; TADAP, translation & adaptation.

*Inferred.

Tool	Validation psychometrics	Cutoffs	Reliability measures	Normative data	Countries
MMSE ◆	Criterion, Construct	Dementia, cognitive decline	Yes	Yes	Egypt, Lebanon, Morocco, Jordan, Tunisia, Qatar, Arabic-speaking[1,2]
MoCA ◆	Criterion, Construct, Content	Dementia, MCI	Yes	Yes	Egypt, Lebanon, Morocco
ACE III ◆	Criterion, Construct	Dementia, MCI	No	Yes	Egypt
Mini-ACE ◆	Criterion, Construct	Dementia	No	No	Egypt
A-TNI93 (GTD-USJ) ◆	Criterion, Construct	Dementia	Yes	Yes	Lebanon
DSB-100 ◆	Construct	No	Yes	No	Tunisia
ADAS-Cog ◆	Criterion, Construct	AD	Yes	Yes	Tunisia
A-RUDAS ◆	Criterion, Construct	Dementia	Yes	No	Lebanon
5WT ◆	Criterion, Construct	AD	No	No	Tunisia
A-IQCODE 16 ◆	Criterion, Construct, Content	Dementia	Yes	No	Lebanon
10/66 ◆	Criterion, Content	Dementia	No	No	Lebanon
ASCA ◆	Construct, Content	No	Yes	No	Egypt
DAS ◆	Criterion, Construct, Content	Dementia	Yes	No	Egypt
AD8 ◆	Criterion, Construct	Dementia	Yes	No	Lebanon
AQ ◆	Criterion, Construct	Dementia	Yes	No	Lebanon
CDR ◆	Criterion, Construct	Dementia	Yes	No	Lebanon
Mini-cog ◆	Criterion, Construct	Dementia	No	No	Qatar
CERAD-ArNB ◆	Construct	No	No	Yes	Oman
DemeGraph ◆	Criterion, Construct	Dementia	No	No	Lebanon
BCST ◆	Construct	No	No	Yes	Arabic-speaking[1]

● Criterion ● Construct ● Content

◆ Size scales with ≥3 settings/study

FIGURE 2 Available psychometric evidence of the validated cognitive assessment tools. 5WT, 5 Words Test; 10/66, 10/66 Dementia Research Group (DRG) diagnostic assessment for dementia; ACE III, Addenbrooke's Cognitive Examination III; AD8, Ascertain Dementia 8-item Informant Questionnaire; ADAS-Cog, Cognitive Subscale of the Alzheimer's Disease Assessment Scale; A-IQCODE 16, Arabic Informant Questionnaire on Cognitive Decline in the Elderly; AQ, the Alzheimer's Questionnaire; A-RUDAS, Arabic Rowland Universal Dementia Assessment Scale; ASCA, Ain Shams Cognitive Assessment Scale; A-TNI93 (GTD-USJ), Arabic Test des Neuf Images du 93-Groupe de Travail sur les Démences de l'Université Saint Joseph; BCST, Brookdale Cognitive Screening Test; CDR, Clinical Dementia Rating scale; CERAD-ArNB, Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Battery-Arabic version; DAS, Dementia Arabic Scale; DSB-100, Dementia Screening Battery-100; Mini ACE, Mini Addenbrooke's Cognitive Examination; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment. [1,2]Arabic speaking in the United States and Israel.

modification of the digit span tests to account for illiteracy).[47] The ASCA was developed as a cognitive battery for mild cognitive impairment (MCI) and dementia diagnosis among Egyptian older adults; the battery development focused on its applicability for illiteracy and high education and on adapting several existing tests (including the Wechsler memory scale, Trail Making Test Part B, digit forward and digit backward test, animal naming, and other naming tests).[48] The DAS scale was constructed for dementia diagnosis among Arabic-speaking populations.[49] The DAS covers several domains (Table 2) and included an informant-based component, and while some information on adaptation is provided, information on battery development and assembly is limited. Translation methods used to develop these tools were unclear, despite them being based on adapting existing cognitive tools.

Psychometric methods and data

We compiled methods and data provided in the studies that validated a tool (Table 2 and a summary in Figure 2). Face/or content validity was tested in six studies for a total of five tools. Construct validity to assess the internal structure of the scales using exploratory factor analysis (EFA) or principal component analysis was only tested in three studies for three tools, while one study performed an EFA for a collection of tools for which normative data were generated.[35] No study used confirmatory factor analysis. Other types of construct validity included external nomological validity ($n = 18$ studies) and known group validity ($n = 12$).

Criterion validity was evaluated in 21 studies and for a total of 16 tools, including eight global cognitive assessments, two domain-specific tests, four informant-based tools, and two cognitive batteries (Table 2; Figure 2). The vast majority involved establishing cut-offs, sensitivity, and specificity for dementia screening (with only three studies investigating MCI and cognitive decline). Studies spanned five countries/populations; Lebanon, Tunisia, Egypt, and Morocco had sensitivity and specificity data for dementia diagnosis and screening for several tools (nine, three, five, and two tools, respectively).

The internal consistency reliability was tested in 17 studies (59%) and for a total of 19 tools while ≈ 40% of the validation studies did not report any measure of internal consistency. Of those that reported reliability measures, eight studies used only one type of reliability measure, six reported two types, while three reported three (Table 2).

Normative data

Normative data were reported in nine studies and for 13 cognitive tools: 12 individual cognitive assessment tools and one battery (CERAD-ArNB). These include seven global cognitive function tests (MMSE, MoCA, ACE III, A-TNI93, ADAS-Cog, the Modified Mini-Mental State Examination [3MS], the BCST), and one battery (CERAD-ArNB). In addition, one study conducted in a Lebanese sample only provided normative data (no validity properties) for five domain-specific tools (the Brief Visuospatial Memory Test-Revised [recall], the Lebanese Digit Span [auditory attention], cross-linguistic naming test [semantic fluency], and the Phonemic and Semantic fluency tests [verbal fluency]).[35] Most of the studies used a sample of healthy individuals to generate the normative data with one exception, the study validating the MMSE in a Tunisian sample, which included MCI and dementia patients.[32]

Performance of cognitive assessment tools

Several tools had high sensitivity and specificity for dementia screening, including the 10/66 battery and several individual cognitive tools (the MoCA, AQ, mini-ACE, the A-IQCODE, and the developed DAS tool). Overall global cognitive assessment performed better than domain-specific tools (5WT and A-TNI93). Although scarce, data on MCI and cognitive decline screening showed poor tool performance, except for the MoCA, which had high sensitivity for MCI screening.

Tools validated in more than one study

We describe in this section the validity, reliability, and the normative data of tools, which were validated in more than one study: the MMSE, MoCA, and ACE III. The MMSE was validated in nine studies in eight countries: Qatar,[22] Lebanon,[35,43] Tunisia,[32] the United States,[34] Israel,[41] Jordan,[42] Morocco,[26] and Egypt.[50] It was validated in clinical and community settings in addition to memory clinics and geriatric clubs. Reported validation properties include criterion validity (against clinical diagnosis, and IQCODE), and construct validity (known-group validity and external nomological validity). Five studies validated the MMSE focusing on dementia screening,[22,26,32,43,50] and different cut-off points were established (ranging from 18 to 26 in five different countries) with a broad range of reported sensitivity (59.1% to 95%) and specificity values (59.5% to 99%). Among those studies, cut-off points for specific groups including illiterate and low-education groups (≤ 9 years' education) were generated in Morocco[26] and Egypt,[50] respectively. One study validated the MMSE in Arab Americans in the United States, and reported a cut-off point of 23 for cognitive decline screening using the IQCODE with a sensitivity of 73% and specificity of 60%.[34] This study also established two different cut-off points, 22 and 23, for low- and high-education groups, respectively. Higher sensitivity and specificity were found with education-based cut-offs and with well-defined diagnostic comparisons.[26,50] Four studies reported normative data for the MMSE: three included cognitively healthy participants (in Tunisa [$n = 73$; age range: 50–95[32]], Jordan [$n = 250$, age range: 50–86[42]], Lebanon [$n = 164$, age range: 60–87[35]]), and among Arabic-speaking older adults living in Israel ($n = 266$, mean age 72.4 ± 5.5[41]). The study conducted in Tunisia also provided normative data for MCI ($n = 57$) and dementia ($n = 56$) patients (age range: 50–95)[32] and the study conducted in Lebanon reported normative data for both MMSE and its modified version (3MS).[35]

MoCA was the second most validated scale, being the subject of five studies in three Arab countries: Lebanon,[35,45] Egypt,[25,37] and Morocco.[24] It was also validated in different settings: clinical, community, memory clinics, and geriatric centers or clubs and validation included criterion validity and construct validity (Table 2). Dementia cut-offs were identified in two studies conducted in Egypt[51] and Morocco,[24] with higher sensitivity and specificity reported in Egypt, potentially explained by the use of severity and education-specific cut-offs.

MCI cut-off points were also reported in a study conducted in Egypt (sensitivity = 92.3%; specificity = 85.7% for a cut-off point of 26[37]). Two studies using the same sample of 164 literate and cognitively healthy Lebanese community-dwelling older adults (age range: 60–87) reported normative data for the MoCA.[35,45] All reported measures of reliability for both MMSE and MoCA were high (> 0.7), demonstrating high internal consistency reliability (Table 2).

The ACE III including the mini version was tested in four studies in Egypt (Table 2), three of which used the same sample.[33,38,39] Dementia cut-off points for ACE III and its shorter version mini-ACE were validated in two studies in Egypt (with high sensitivity and specificity).[33,38] Another study using the same sample reported ACE III cut-off of 81 for MCI with a 75% sensitivity and 82% specificity.[39] Normative data for ACE III were reported in one study among a subsample of 56 cognitively healthy Egyptian older adults (age range: 60–93). No reliability measures were available for this test. All other cognitive assessment tests were validated in only one study or setting as presented in Tables 1 and 2.

Reported influencing factors

The most reported factor influencing test performance was education, investigated in 18 of the 29 validity studies (62%; Table 1), with most reporting that low education groups performed poorer on cognitive tests,[24–27,30,32,34–36,41,42,44,45,47,48,50] or even could not perform the test for some.[22] Among these, five reported education-based cut-off points.[25,26,34,49,50] Age was investigated in 13 studies reporting poorer performance among older adults.[22,24,30,32,35,36,40–42,44–47] Other factors specific to each study are included in Table 1. Sex was investigated in four studies, with mixed results.[35,36,41,46] Some studies merely mentioned the potential influence of education (n = 5), age (n = 2), and cultural factors (n = 4) but did not formally investigate them.

Cognitive assessment tools used in studies

The characteristics of the 125 studies using a cognitive assessment tool are presented in supporting information Table S1. The included studies were published between 1998 and 2023, and 56 (45%) of the studies were published in the past 5 years. Twenty-one countries were represented with most studies being conducted in Egypt (n = 28), Kingdom of Saudi Arabia (n = 26), Lebanon (n = 20), Israel (n = 14), Morocco (n = 10), Jordan (n = 9), and Tunisia (n = 9). Aside from a case report, the sample size ranged between 25 and 193,715. Only 36 out of the 125 included studies reported using a validated Arabic version of the instruments used. In addition, seven studies used multiple assessment tools, but they only reported the use of a validated Arabic version for one of those tools. Although involving Arabic-speaking participants, 48 studies did not explicitly state that the cognitive tests were administered in Arabic; hence, the language in which the tool was administered remained unclear. Furthermore, three studies used more than one tool but only mentioned the language by which one of the tools was administered and provided no information regarding the other tools.

The included studies used 51 different cognitive evaluation tools/instruments (Table S1). Seven studies did not clarify the name of at least one of the tests administered, three of which used the same sample and tools.[52–54] The MMSE (including the modified version) was the most frequently used instrument, appearing in 78 of the studies with 29 referring to using an Arabic validated version of the tool. Nineteen studies used the MoCA, making it the second most popular instrument, with 16 reporting using an Arabic-validated MoCA.

4 | DISCUSSION

This review comprehensively examined studies that have validated cognitive assessment tools among the Arabic-speaking older adult population in the Arab world and beyond. A total of 29 studies were identified, validating 20 cognitive assessment tools in this population. Cut-off points for dementia screening or diagnosis were validated for 16 cognitive tools, the majority being global cognitive assessment tools and batteries, with only two validated domain-specific cognitive tools. Only three tools—MMSE, MoCA, and ACE III—were evaluated for MCI and cognitive decline detection, highlighting critical limitations in available assessments for earlier cognitive changes. These findings underscore important limitations regarding validated and robust cognitive assessments for Arabic-speaking populations, posing significant challenges for addressing the growing ADRD burden in this population.[1,4]

Moreover, studies were conducted in a limited number of settings and countries. Most of the validated tools were evaluated in clinical samples, limiting their applicability and generalizability for diverse or community samples, and only a few were tested in age- or education-specific subgroups. These are important limitations given the large variability in language, education levels, and cultural factors across Arabic-speaking populations.

Although Arabic is the official language across Arab nations, and MSA is used in education, administration, and media, each region has its own colloquial Arabic dialect (local Arabic). Variations and differences in these dialects can range from differences in pronunciation to vocabulary and structural changes, influenced by factors such as regional contexts, rural and urban settings, historical influences, and cultural factors. Moreover, certain terms cannot be translated with a single equivalent word, and some concepts may lack direct cultural translation.[55] Research also suggests that performance on cognitive tests can vary depending on the cultural importance of the specific cognitive skill being assessed.[16] These challenges underscore the need for thorough examination and adaptation of cognitive tests for the heterogeneous Arabic-speaking population. Yet, based on our results, cognitive assessments have been investigated in 7 out of 22 Arab countries and only three countries had more than one validated tool, leaving several nations and regions (including the Gulf countries, which have a distinct dialect, and large-population countries such as Algeria, Iraq, and Kingdom of Saudi Arabia) with no validated cognitive assessments in their populations.

Our review also showed that only 11 out of 29 validation studies reported using colloquial Arabic when validating the cognitive assessment tools, whereas others used formal Arabic or not clearly specified forms of Arabic. This poses additional difficulties in understanding the applicability and transferability of the tool across Arabic-speaking communities. Moreover, the use of formal Arabic can introduce important challenges as this form of Arabic is more related to academic and studied Arabic, making the cognitive tests more prone to being influenced by the education level of participants.

Another complexity arises from the wide diversity in this population, which spans high rates of illiteracy and of multilingualism. Arab countries have some of the highest rates of illiteracy and low literacy,[56] and several of the reviewed studies focused on tool adaptations and development to capture cognitive outcomes in this higher -risk sub-group.[22,26,46–49] In parallel, several Arab countries (such as Morocco and Lebanon) have a high percentage of bilingualism, and several languages are commonly used in educational and official settings, with many higher education programs taught in languages other than Arabic.[57,58] This language–education interplay emphasizes the importance of tailoring tests based on populations' realities and contexts.[59,60] As an illustration, we note the reported observations by Hayek et al., in their MoCA validation study in a sample of Lebanese older adults, that participants struggled with the use of formal Arabic in the fluency domain; the authors attributed this to Lebanon's multilingual environment and reported that participants responded to the fluency task by providing examples in other languages.[45] In contrast, Albanna et al. reported that illiteracy hindered completion of the mini-cog cognitive assessment in Qatar.[22] We also note that education was an important factor influencing cognitive performance in the validation studies reviewed, and that the use of education-based cut-offs improved psychometric properties of validated tools, further underscoring the value of more comprehensive validation efforts that account for sociodemographic-, education-, and population-specific contexts.

Another major finding was that reported sensitivity and specificity were high and promising for several individual tools, cognitive batteries, and informant-based tests. More data are needed to reproduce and confirm these findings, particularly given the limited settings and sample size of the validation studies. The MMSE, MoCA, and ACE III were the only tools validated in different studies with the MMSE and MoCA validated in more than one country. The MOCA showed more consistent sensitivity and specificity, whereas the MMSE had more varying psychometrics across the studies (sensitivity ranging from 59.1% to 95% and specificity from 59.5% to 99%). This variability could be explained by different factors. The MMSE was the most validated tool among older adults in Arab countries, which could generate more variability in results. This variability could also be potentially reflecting the variability in the validation methods used. MMSE sensitivity and specificity were low in studies with less clearly established diagnostic comparisons[34,43]; for example, using another tool (such as the IQCODE) as the reference test could introduce measurement error and difficulties given the lack of well-validated scales in this population. Another explanation could be the variability in study sample and settings; for example, the low sensitivity and specificity reported in Albanna et al. was attributed to the higher illiteracy rate and older age of the sample.[22] It could also be explained by the use of formal Arabic in a mixed population (including Qataris and non-Qataris).

Different MMSE cut-off points were established across studies, with higher cut-off scores reported in younger and higher education samples.[26,32,50] In addition, all the studies that assessed the validity of MMSE reported that it was affected by education (Table 1). Together these findings emphasize the importance of establishing population- and subgroup-specific cut-offs. They also underline the importance of robust methods and high-quality reporting for an accurate characterization of tool performance and its variability. In our review, we found a large heterogeneity in the validity and reliability methods used across studies, with 21% of the validation studies reporting only one form of validity and 41% of the studies reporting any reliability measure (and none conducting confirmatory factor analysis). As indicated in Table 2, we had to infer at least one of the validation methods used in 23 of the 29 validation studies, further emphasizing the need for rigorous methodology and reporting.

Another important gap identified in the review concerns the validation of domain-specific tools and the development of cognitive batteries. Research efforts on improving cognitive assessment, notably initiatives like HCAP, have been prioritizing the use of a variety of validated domain-specific tools and a comprehensive battery assembly to better characterize early changes across multiple areas of cognition leading to cognitive decline/dementia.[8] The limited efforts and data on domain-specific tools and combined batteries pose significant limitations for the implementation of such approaches in Arabic-speaking populations, hindering opportunities for more comprehensive cognitive assessments within a population as well as cross-population harmonization and comparisons. Information on normative data was also limited, with available norms based on small samples and not accounting for key demographic and socioeconomic factors, emphasizing the need for improved characterization of normative cognitive data across Arabic-speaking older communities.

In terms of cognitive tools used, the MMSE and MoCA were the most widely used cognitive tools among Arabic-speaking older adults (being used in > 60% of the published studies on cognitive functioning). Given the education, language, and cultural heterogeneity of Arabic-speaking countries, this raises questions regarding the robustness of reported cognitive data. For example, given their high correlation with education, using the MMSE or MoCA could overestimate the prevalence of cognitive impairment in low-literacy settings.[13]

To our knowledge, this is the first systematic review that collates a comprehensive list of all available validated cognitive assessment tools for Arabic-speaking older adults and examines their psychometric properties and use. We used a sensitive and thorough search strategy, identifying studies that validated and used a cognitive assessment tool among Arabic-speaking older adults, to present a broad overview on cognitive assessments and cognitive data. Both the selection and data abstraction steps were done in duplicate by two independent reviewers. The review includes several limitations. First, the small number of validation studies made it difficult for us to draw firm conclusions about the best validated cognitive assessment instruments and their optimal cut-off points for different ages, sexes, and education level groups. Second, the heterogeneity of the included studies and some suboptimal reporting led us to have to infer the validation methods used. Given these challenges, we were not able to conduct systematic quality assessments for all studies.

5 | CONCLUSION

This review identified major gaps regarding the development, adaptation, and validation of cognitive assessments for Arabic-speaking older adults. Overall, our findings are in line with literature reporting that dementia research productivity is low in the MENA region[61,62] despite elevated dementia prevalence.[4] Dementia research and policies will continue to be hindered in the absence of robustly validated and tested cognitive assessment tools and batteries in this population. Recent progress in validating widely used global cognitive assessment tools in Arabic-speaking older adults is promising, yet important gaps need to be addressed to effectively advance ADRD research, prevention, and management in Arab populations. Our findings identify critical needs for high-quality data on psychometric properties of cognitive assessments and for expanding both the tools/cognitive domains and populations and at-risk subgroups studied. First, there is a large need for validation efforts across different nations and populations. These efforts can benefit from building on tool adaptations performed in samples with comparable dialects and cultural norms. Second, reproducibility studies are needed for validated tools. Both validation and reproducibility studies should prioritize implementing large and representative sampling and robust methodology: sample size estimation, the incorporation of multiple validity methods, including criterion validity evaluations with well-defined and assessed reference test, and confirmatory factor analysis for rigorous assessment of the construct validity of cognitive tools. A high priority should be given to generating education-, age-, and sex-specific psychometrics and cut-offs and to carefully examining relevant sociocultural factors such as low literacy and multilingualism. Relatedly, high-quality reporting is instrumental to build much needed evidence in an efficient and impactful manner. This concerns validity and reliability reporting as well as thorough reporting of tool and battery adaptation and development. Another priority for future studies is the validation of domain-specific cognitive tools and their assembly into cognitive batteries that give more comprehensive assessments of several cognitive domains and abilities. Finally, our results underscore the critical need for validating assessments for MCI, cognitive decline, and cognitive changes, which are instrumental for advancing early detection, prevention, and risk factor identification.

AUTHOR CONTRIBUTIONS

Mayssan Kabalan: Conceptualization; protocol writing and registration; search strategy; screening and data abstraction; data analysis and interpretation; manuscript writing; revising; and approval of the final version. **Ola Bazzi**: Search strategy; screening and data abstraction; data analysis and interpretation; and revision and approval of the final version of the manuscript. **Ola El Zein**: Search strategy and revision and approval of the final version of the manuscript. **Josleen Al Barathie and Lara Chehabeddine**: Data abstraction and revision and approval of the final version of the manuscript. **Joanne Khabsa**: Protocol writing; search strategy; and revision and approval of the final version of the manuscript. **Monique Chaaya and Carlos Mendes de Leon**: Conceptualization; protocol writing; search strategy; and revision and approval of the final version of the manuscript. **Martine Elbejjani**: Conceptualization; protocol writing; search strategy; data analysis and interpretation; manuscript writing; and revision and approval of the final version of the manuscript.

ACKNOWLEDGMENTS

This systematic review is supported by National Institutes of Health under the grant number: 1R01 AG069016-01. Title: Dementia and Related Health and Social challenges in Lebanon: A Population Study. The funder did not play any role in study design, analysis, interpretation of data, or the writing and submission of the report.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. Author disclosures are available in the supporting information.

CONSENT STATEMENT

This article does not contain any data from human participants or animals, hence informed consent was not required.

ORCID

Martine Elbejjani https://orcid.org/0000-0002-5035-0370

REFERENCES

1. World Health Organization. Alzheimer disease and other dementias. 2013. https://www.medbox.org/dl/5e148832db60a2044c2d5489
2. World Health Organization. Dementia 2021. https://www.who.int/news-room/fact-sheets/detail/dementia
3. Alzheimer's Disease International. *The Global Impact of Dementia 2013-2050*. Alzheimer's Disease International; 2013. https://www.alzint.org/u/2020/08/GlobalImpactDementia2013.pdf
4. El-Metwally A, Toivola P, Al-Rashidi M, et al. Epidemiology of Alzheimer's disease and dementia in Arab countries: a systematic review. *Behav Neurol*. 2019;2019:3935943.
5. Manly JJ, Jones RN, Langa KM, et al. Estimating the prevalence of dementia and mild cognitive impairment in the US: the 2016 health and retirement study Harmonized Cognitive Assessment Protocol project. *JAMA Neurol*. 2022;79(12):1242-1249.
6. Kobayashi LC, Jones RN, Briceño EM, et al. Cross-national comparisons of later-life cognitive function using data from the Harmonized Cognitive Assessment Protocol (HCAP): considerations and recommended best practices. *Alzheimers Dement*. 2024;20(3):2273-2281.
7. Gross AL, Li C, Briceño EM, et al. Harmonisation of later-life cognitive function across national contexts: results from the Harmonized Cognitive Assessment Protocols. *Lancet Healthy Longev*. 2023;4(10):e573-e583.
8. Langa K, Ryan L, McCammon R, et al. The health and retirement study Harmonized Cognitive Assessment Protocol project: study design and methods. *Neuroepidemiology*. 2020;54(1):64-74.
9. Chan KS, Gross AL, Pezzin LE, Brandt J, Kasper JD. Harmonizing measures of cognitive performance across international surveys of aging using item response theory. *J Aging Health*. 2015;27(8):1392-1414.
10. Briceño EM, Arce Rentería M, Gross AL, et al. A cultural neuropsychological approach to harmonization of cognitive data across culturally and linguistically diverse older adult populations. *Neuropsychology*. 2023;37(3):247-257.
11. Kobayashi LC, Gross AL, Gibbons L, et al. You say tomato, i say radish: can brief cognitive assessments in the U.S. Health Retirement Study be harmonized with its international partner studies? *J Gerontol B Psychol Sci Soc Sci*. 2021;76(9):1767-1776.
12. Rosli R, Tan MP, Gray WK, Subramanian P, Chin A. Cognitive assessment tools in Asia: a systematic review. *Int Psychogeriatr*. 2016;28(2):189-210.
13. Arévalo S, Kress J, Rodriguez F. Validity of cognitive assessment tools for older adult hispanics: a systematic review. *J Am Geriatr Soc*. 2020;68(4):882-888.
14. Harb C. The Arab region: cultures, values, and identities. *Handbook of Arab American psychology*. 2016:3-18.
15. World Bank. Population, total—Arab World. 2023. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=1A
16. Ostrosky-Solis F, Ramirez M, Ardila A. Effects of culture and education on neuropsychological testing: a preliminary study with indigenous and nonindigenous population. *Appl Neuropsychol*. 2004;11(4):186-193.
17. Alzheimer's Disease International. World Alzheimer Report. Improving health care for people living with dementia; coverage, quality and costs now and in the future. 2016. https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf
18. Page MJ, McKenzie JE, Bossuyt P, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
19. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.

20. Bartram D, Berberoglu G, Grégoire J, Hambleton R, Muñiz J, van de Vijver F. *ITC guidelines for translating and adapting tests* (2nd ed.). Int J Test. 2018;18(2):101-134.
21. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.
22. Albanna M, Yehya A, Khairi A, et al. Validation and cultural adaptation of the Arabic versions of the Mini-Mental Status Examination-2 and Mini-Cog test. *Neuropsychiatr Dis Treat*. 2017;13:793-801.
23. Phung KTT, Chaaya M, Waldemar G, et al. Validation of the 10/66 Dementia Research Group diagnostic assessment for dementia in Arabic: a study in Lebanon. *J Geriatr Psychiatry Neurol*. 2014;27(4):282-290.
24. Khatib N, El Harch I, Lamkaddem A, et al. The Moroccan MoCA test: translation, cultural adaptation, and validation. *Appl Neuropsychol Adult*. 2022:1-5.
25. Saleh AA, Alkholy RSAEHA, Khalaf OO, et al. Validation of Montreal Cognitive Assessment-Basic in a sample of elderly Egyptians with neurocognitive disorders. *Aging Ment Health*. 2019;23(5):551-557.
26. Rami Y, Diouny S, Kissani N, Yeou M. Cross-cultural adaptation of the Moroccan version of the Mini-Mental State Examination: a preliminary study. *Appl Neuropsychol Adult*. 2022:1-6.
27. Chaaya M, Phung T, El Asmar K, et al. Validation of the Arabic Rowland Universal Dementia Assessment Scale (A-RUDAS) in elderly with mild and moderate dementia. *Aging Ment Health*. 2016;20(8):880-887.
28. Phung TKT, Chaaya M, Asmar K, et al. Performance of the 16-Item Informant Questionnaire on Cognitive Decline for the Elderly (IQCODE) in an Arabic-speaking older population. *Dement Geriatr Cogn Disord*. 2015;40(5-6):276-289.
29. Nielsen TR, Phung TKT, Chaaya M, Mackinnon A, Waldemar G. Combining the Rowland universal dementia assessment scale and the informant questionnaire on cognitive decline in the elderly to improve detection of dementia in an Arabic-speaking population. *Dement Geriatr Cogn Disord*. 2016;41(1-2):46-54.
30. Ben Jema S, Attia R, Bahri-Mrabet A, et al. An Arabic Version of the Cognitive Subscale of the Alzheimer's Disease Assessment Scale (ADAS-Cog): reliability, validity, and normative data. *J Alzheimers Dis*. 2017;60(1):11-21.
31. Karam G, Itani T, Nasreddine Z, et al. Validation into Arabic versions of dementia rating scales, dementia caregivers scales, and dementia research instruments. *Alzheimers Dement (Amst)*. 2018;10:796-801.
32. Bellaj T, Ben Jemaa S, Rachidi L, et al. Version arabe du mini mental state examination (A-MMSE): fidélité, Validité et données normatives. *La Tunisie médicale*. 2008;88:768-776.
33. Qassem T, Khater MS, Emara T, et al. Adaptation and validation of the mini-Addenbrooke's cognitive examination in dementia in Arabic speakers in Egypt. *Dement Geriatr Cogn Disord*. 2020;49(6):611-616.
34. Wrobel NH, Farrag MF. Preliminary validation of an Arabic version of the MMSE in the elderly. *Clinical Gerontologist*. 2008;31:75-93.
35. Abou-Mrad F, Chelune G, Zamrini E, Tarabey L, Hayek M, Fadel P. Screening for dementia in Arabic: normative data from an elderly Lebanese sample. *Clin Neuropsychol*. 2017;31(1):1-19.
36. Alobaidy A, Al-Sinawi H, Al-Asmi A, et al. CERAD neuropsychological battery – Arabic version: regression-based and stratified normative data and effects of demographic variables on cognitive performance in older Omanis. *Cogn Behav Neurol*. 2017;30:150-158.
37. Rahman T, El Gaafary M. Montreal Cognitive Assessment Arabic version: reliability and validity prevalence of mild cognitive impairment among elderly attending geriatric clubs in Cairo. *Geriatr Gerontol Int*. 2009;9(1):54-61.
38. Qassem T, Khater M, Emara T, et al. Validation of the Egyptian-Arabic Version of the Addenbrooke's Cognitive Examination III (ACE-III) in diagnosing dementia. *Dement Geriatr Cogn Disord*. 2020;49(2):179-184.
39. Qassem T, Khater MS, Emara T, et al. Validation of the Addenbrooke's Cognitive Examination-III in mild cognitive impairment in Arabic speakers in Egypt. *Dement Geriatr Cogn Disord*. 2020;49(4):418-422.
40. Qassem T, Khater MS, Emara T, et al. Normative data for healthy adult performance on the Egyptian–Arabic Addenbrooke's Cognitive Examination III. *Middle East Curr Psychiatry*. 2015;22:27-36.
41. Inzelberg R, Schechtman E, Abuful A, et al. Education effects on cognitive function in a healthy aged Arab population. *Int Psychogeriatr*. 2007;19(3):593-603.
42. Dahbour S, Hashim M, Alhyasat A, et al. Mini-mental state examination (MMSE) scores in elderly Jordanian population. *Cereb Circ Cogn Behav*. 2021;2:100016.
43. El-Hayeck R, Baddoura R, Wehbé A, et al. An Arabic version of the Mini-Mental State Examination for the Lebanese population: reliability, validity, and normative data. *J Alzheimers Dis*. 2019;71(2):525-540.
44. Khiari Mrabet H, Attia N, Bellaj T, Bennys K, Anane N, Mrabet A. Version arabe de L'épreuve des 5 mots: Validation clinique pour le diagnostic de démence de type alzheimer. *La Tunisie médicale*. 2008;88:786-792.
45. Hayek M, Tarabey L, Abou-Mrad T, Fadel P, Abou-Mrad F. Normative data for the montreal cognitive assessment in a Lebanese older adult population. *J Clin Neurosci*. 2020;74:81-86.
46. El-Hayeck R, Baddoura R, Wehbé A, et al. An adapted Arabic version of the test of nine images for the illiterate Lebanese population: validation and preliminary normative data. *J Int Neuropsychol Soc*. 2023;29(3):316-323.
47. Bellaj T, Ben Jemaa S, Khelifa M, Ben Djebara M, Gouider R, Le Gall D. The development of the dementia screening battery-100: instrument presentation, reliability, and construct validity. *Dement Geriatr Cogn Dis Extra*. 2017;7(2):215-229.
48. Elkholy N, Tawfik HM, Ebeid S. A model of cognitive evaluation battery for diagnosis of mild cognitive impairment and dementia in educated and illiterate Egyptian elderly people. *Egypt J Neurol Psychiatr Neurosurg*. 2020;95:1-7.
49. Farghaly WM, El Tallawy HN, Mohamed H, El Tallawy SH, Abdelhamed MA. Construction and standardization of dementia arabic scale. *Neuropsychiatr Dis Treat*. 2021;17:721-729.
50. ElKholy N, Tawfik HM, Ebeid S, et al. Defining cut-off scores for MMSE in an educated and illiterate Arabic speaking Egyptian elderly population. *Egypt J Geriatr Gerontol*. 2018;6:1-3.
51. Saleh AA, Alkholy RSAEHA, Khalaf OO, et al. Validation of Montreal Cognitive Assessment-Basic in a sample of elderly Egyptians with neurocognitive disorders. *Aging Ment Health*. 2019;23(5):551-557.
52. El Kadmiri N, Zaid N, Hachem A, et al. Novel mutations in the amyloid precursor protein gene within Moroccan patients with Alzheimer's disease. *J Mol Neurosci*. 2014;53(2):189-195.
53. El Kadmiri N, Zaid N, Zaid Y, et al. Novel presenilin mutations within Moroccan patients with Early-Onset Alzheimer's Disease. *Neuroscience*. 2014;269:215-222.
54. Kadmiri N, Younes Z, Hamzi K, Nadifi S, Slassi I, El Moutawakil B. Présentation clinique de cas marocains atteints de la maladie d'Alzheimer. *L'Encéphale*. 2014;40:40.
55. Abou-Mrad F, Tarabey L, Zamrini E, et al. Sociolinguistic reflection on neuropsychological assessment: an insight into selected culturally adapted battery of Lebanese Arabic cognitive testing. *Neurol Sci*. 2015;36(10):1813-1822.
56. Arab League Educational, Cultural and Scientific Organization (ALECSO). ALECSO Observatory: The number of illiterates in the Arab world could reach 100 million by 2030. 2023. www.alecso.org
57. Rodriguez Farah M, Lugo E, Bachioqui EL. Bilingual education in the MENA region – An overview. *Plurilinkgua*. 2020;16(2):1–12.
58. Zakharia Z, Education M, Garcia O, Lin A, May S. Bilingual education in the Middle East and North Africa. In: García O, Lin A, May S, eds. *Bilingual and Multilingual Education*. Springer International Publishing; 2016:1-13.

59. Ardila A. Language representation and working memory with bilinguals. *J Commun Disord*. 2003;36(3):233-240.
60. Abdelgafar G, Moawad R. Executive function differences between bilingual Arabic–English and monolingual Arabic children. *J Psycholinguist Res*. 2015;44:651-667.
61. Al Sinawi, H., Qassem T, Al Harrasi A, et al. In: Laher, I., ed. *Handbook of Healthcare in the Arab World*. Springer International Publishing; 2021:2953-2966.
62. Hatahet O, Roser F, Seghier M. Cognitive decline assessment in speakers of understudied languages. *Alzheimers Dement (N Y)*. 2023;9(4):e12432.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kabalan M, Bazzi O, Al-Barathie J, et al. Cognitive assessment tools for Arabic-speaking older adults: A systematic review. *Alzheimer's Dement*. 2025;21:e70207. https://doi.org/10.1002/alz.70207