Research Article

Do normative data specific to Greek Australian older adults improve validity of neuropsychological assessment results?

Mathew Staios[1,7] ◉, Mary H. Kosmidis[2] ◉, Yiannis Tsiaras[2] ◉, T. Rune Nielsen[3] ◉, Alexandra Papadopoulos[1],
Arthur Kokkinias[4], Dennis Velakoulis[5,6], Evrim March[1] and Renerus J. Stolwyk[1,7] ◉

[1]Turner Institute for Brain and Mental Health, Monash University, Melbourne, Australia, [2]Lab of Cognitive Neuroscience, School of Psychology, Aristotle University of Thessaloniki, Thessaloniki, Greece, [3]Department of Neurology, Neuroscience Centre, Danish Dementia Research Centre, Copenhagen University Hospital, Copenhagen, Denmark, [4]The Royal Melbourne Hospital, Inner West Area Mental Health Service, Melbourne, Australia, [5]Department of Psychiatry, Melbourne Neuropsychiatry Centre, The University of Melbourne and Melbourne Health, Melbourne, Australia, [6]Neuropsychiatry, The Royal Melbourne Hospital, Parkville, Melbourne, Australia and [7]Monash-Epworth Rehabilitation Research Centre, Melbourne, Australia

Abstract

Objective: This study aimed to compare Greek Australian and English language normative data with regard to impairment rates yielded within a healthy Greek Australian older adult sample. We also examined whether optimal cut scores could be identified and capable of sensitively and specifically distinguishing between healthy Greek Australians from those with a diagnosis of Alzheimer's disease (AD). **Method:** Ninety healthy Greek Australian older adults and 20 demographically matched individuals with a diagnosis of AD completed a range of neuropsychological measures, including the Wechsler Adult Intelligence Scale-Fourth Edition, Greek Adaptation (WAIS-IV GR), verbal and visual memory, language and naming, and executive functions. Impairment rates derived from the use of either Greek Australian or English language normative data were calculated and compared, using a 1.5 standard deviation criterion to denote impairment. Receiver operating characteristics curve analysis was used to investigate the sensitivity and specificity of alternate cut scores. **Results:** Impairment rates derived from the Greek Australian normative data showed that rates of impairment generally fell within the expected 7% range. In contrast, impairment rates for all tests derived using English language normative data were significantly higher and ranged from 11%–66%. Comparisons between healthy and AD participants with moderate dementia showed significant differences across all measures. Area under the curve results ranged from .721 to .999 across all measures, with most tests displaying excellent sensitivity and specificity. **Conclusions:** English language normative data were found to be inappropriate for use with Greek Australian elders, potentially leading to erroneous diagnostic outcomes. The use of minority group specific normative data and associated cut points appear to partially ameliorate this issue. Clinical implications are discussed alongside future research directions.

Keywords: Greek Australian; Alzheimer's disease; neuropsychological assessment; diagnostic accuracy; cross-cultural neuropsychology

(Received 17 January 2023; final revision 21 July 2023; accepted 25 July 2023)

As of 2020, there were over 55 million people worldwide living with dementia. This number is expected to almost double every 20 years, reaching approximately 139 million by 2050 (World Health Organisation, 2022). Projected trends suggest that over 700,000 Australian older adults are expected to meet criteria for dementia by 2050, with one-third consisting of individuals from culturally and linguistically diverse backgrounds (Australian Institute of Health & Welfare, 2018). Research has consistently reported that many English language neuropsychological measures and normative data are prone to misclassification when applied to culturally and linguistically diverse groups (Daugherty et al., 2017; Heaton et al., 2003). Poor test specificity has been attributed to cultural and linguistic heterogeneity, level and quality of education, use of unrepresentative normative data, culturally biased test content, and test-taking attitudes (Rivera Mindt et al., 2010;

Shuttleworth-Edwards, 2016). These issues have been extensively examined within minority groups in the United States, including African Americans and Hispanics, where healthy individuals were over three times more likely to be misclassified as impaired (Heaton et al., 2003). In response to this issue, demographically focused normative data have been developed across a range of culturally diverse populations (Nielsen et al., 2018; Pienaar et al., 2016). While it is anticipated that normative data specifically developed for ethnic minority groups will improve the validity of neuropsychological assessment, surprisingly limited empirical research has examined this issue. The Neuropsychological Norms for the U.S.-Mexico Border Region in Spanish Project (NP-NUMBRS) reported that their adapted tests and normative data more accurately classified healthy Spanish-speaking individuals and yielded expected rates of

impairment for a majority of measures compared to published normative data for English-speaking non-Hispanic Whites and Blacks in the United States (Díaz-Santos et al., 2021; Marquine et al., 2020; Marquine et al., 2021; Scott et al., 2021; Suarez et al., 2021). However, further research across multiple international contexts is required to ascertain whether the use of adapted tests and minority group reference normative data indeed results in reducing misclassification rates in healthy culturally diverse individuals and improves diagnostic accuracy across a range of neurocognitive disorders.

In the absence of valid and reliable neuropsychological measures, accurate assessment of cognitive disorders within ethnic minority and/or non-English-speaking groups continues to be an ongoing challenge facing the global neuropsychological community, including Australia (Blakemore et al., 2018; Franzen et al., 2021; Rivera Mindt et al., 2010; Wallace et al., 2018). Similar to other developed nations, illiteracy rates in Australia are very low and the average level of education is high (Australian Institute of Health and Welfare, 2021). In contrast, a large proportion of aging Australian ethnic minorities attained low levels and a poorer quality education prior to immigrating, relative to standards noted in their Australian-born elderly counterparts (Fratti et al., 2011; Jupp, 2001; Noutsos, 2003; Plitas et al., 2009). A case in point is elderly Greek Australian immigrants, who represent one of the largest aging ethnic minority groups in Australia, totaling approximately 63,000 individuals aged 65 years and older, with over 40,000 having attained a primary level of education or less (Australian Bureau of Statistic, 2016; Fanany & Avgoulas, 2019; Staios, 2022). Accordingly, performance on most neuropsychological tests is heavily influenced by years (and quality) of education (Shuttleworth-Edwards, 2016). However, older low educated immigrant groups are underrepresented in a majority of normative studies, making it extremely difficult to interpret what constitutes a normal performance on several neuropsychological measures (Kosmidis, 2017; Nielsen & Waldemar, 2021). Given the noted limitations with using English language tests and normative data, we adapted test content and developed normative data for a broad range of neuropsychological measures, including general intelligence (i.e., Wechsler Adult Intelligence Scale-Fourth Edition [WAIS-IV] Greek Adaptation; Wechsler, 2014), verbal and visual memory tests, language and naming, and executive functioning measures for use with Greek Australian older adults (Staios et al., 2023a, b). The first aim of this study was to compare our newly established Greek Australian normative data with existing published English language normative data in terms of impairment rates they yield within a healthy Greek Australian older adult sample. We hypothesized that using Greek Australian normative data would produce lower impairment rates, while English language normative data would produce significantly higher rates of impairment. The second aim of the study was to examine whether Greek Australian normative data could yield sensitive and specific cut scores to distinguish healthy Greek Australians from Greek Australians with a diagnosis of Alzheimer's disease (AD).

Method

Participants

Healthy participants were recruited using convenience sampling from several Greek social clubs throughout the Melbourne metropolitan area. In order to be eligible for participation, participants had to be aged between 70–85 years, literate, and immigrants from Greece, with Greek as their dominant language. The exclusion criteria for the participants were: (i) a score on the Mini-Mental State Examination below 22/30 (MMSE; Fountoulakis et al., 2000; Plitas et al., 2009); (ii) a score of < 0.5 on the Clinical Dementia Rating Scale (CDR; Hughes et al., 1982); (iii) history of serious neurological or psychiatric conditions known to impact cognition; (iv) a score ≥ 6/15 on the Geriatric Depression Scale-15 (GDS-15) (Fountoulakis et al., 1999) or a score of ≥ 8/20 on the Geriatric Anxiety Inventory (GAI; Pachana et al., 2007); (v) evidence of long term alcohol or substance abuse; and (vi) uncorrected sensory and/or motor deficits.

Participants with AD were recruited via specialist outpatient clinics throughout the Melbourne metropolitan area. Inclusion and exclusion criteria were the same as for healthy participants but also requiring a probable diagnosis of AD (therefore MMSE, CDR exclusion criteria did not apply). The clinical diagnosis of AD was classified using the following methods: interview with the patient and (when possible) an informant; a neurological, physical, and psychiatric examination; cognitive screening with the Greek version of the MMSE or the Neuropsychiatry Unit Cognitive Assessment Tool (NUCOG; Walterfang et al., 2003); a score of ≥1 on the CDR; laboratory screening with blood tests, and structural brain imaging with computerized tomography, magnetic resonance imaging, and/or functional brain imaging with single-photon emission computed tomography. Diagnoses were based on using evidence from all clinical results, in conjunction with the established diagnostic criteria as described by the Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition (DSM-5; American Psychiatric Association, 2013) for neurocognitive disorders and the diagnostic research criteria for AD (Dubois et al., 2007).

Measures

The measures presented in the section to follow were previously administered in the Greek language and normed for use with Greek Australian older adults. This has been described in detail elsewhere (Staios et al., 2023a, b). To compare performances between normative data sets, we selected comparable Australian tests and normative data routinely used in clinical practice for the assessment of older adults. In the absence of local Australian normative studies, we employed tests and normative data from the United States and Canada, which are also routinely used for the assessment of older adults in Australia. For the present study, the term English language normative data will be used throughout this paper when referring to existing published normative data developed for the majority English-speaking population, including normative data published in test manuals and research papers.

Cognitive screener

Mini-Mental State Exam – Greek Adaptation (MMSE; Fountoulakis et al., 2000). Previous research has established that using the recommended Greek national MMSE cutoff of < 24/30 has resulted in large proportions of healthy Greek Australian older adults falling within the impaired range (i.e., 35% scored below the MMSE cutoff score of < 24/30; Plitas et al., 2009). Therefore, we selected an MMSE cutoff of < 22/30, which has been recommended when screening low educated ethnic minority populations (Kochhann et al., 2010; Pedraza et al., 2011).

General intelligence

The Wechsler Adult Intelligence Scale-Fourth Edition, Greek Adaptation (WAIS-IV GR; Wechsler, 2014). The Greek version of the WAIS-IV was adapted from the original version of the WAIS-IV (Wechsler, 2008) and includes the same subtests, items, and score structure with only minor cultural adaptations in item content/scoring. It includes 10 core and five supplemental subtests and yields a full-scale IQ (FSIQ) and four index scores (verbal comprehension index, VCI; perceptual reasoning index, PRI; working memory index, WMI, and processing speed index, PSI). English language normative data used to calculate standard performances were derived using US data (Wechsler, 2008).

New learning & memory

The Hopkins Verbal Learning Test – Revised Form 5 (HVLT-R; Brandt & Benedict, 2001). The HVLT-R is a verbal memory task consisting of 12 words from three semantic categories (professions, food, and sports) presented over three trials, with correct words summed to equal a total score (range = 0–36). Following a 30-min interference interval, participants are required to recall as many words learned from the list presented (range = 0–12), followed by a yes/no recognition trial. The recognition component has a total of 24 words: 12 target words, 6 semantically related words, and 6 semantically unrelated words. English language normative data used to calculate standard performances were derived using Australian data (Hester et al., 2004).

Greek Story Memory Tests (Kosmidis et al., 2012). The Greek Story Memory Tests assesses narrative memory under a free recall condition. A short story relating to a couple taking a holiday on a Greek island is presented over two trials, with immediate recall after each trial (range = 0–32). After a 30-min interference interval, participants are required to recall as much of the story as possible (range = 0–16), followed by a 10-question recognition trial (range = 0–10).

Wechsler Memory Scale - Fourth Addition (WMS-IV; Wechsler, 2009) Visual Reproduction. WMS-IV Visual Reproduction assesses memory for visual stimuli. Five pages of designs are presented in turn, for 10 s each. After each page is presented, the examinee is asked to draw the designs (range = 0–43). After a 30-min interference interval, participants are asked to draw as many designs as they can remember, in any order (range = 0–43). Finally, a recognition trial is administered where the examinee is required to select the target design among six options (range = 0–7). English language normative data used to calculate standard performances were derived using US data (Wechsler, 2009).

Language & naming

Greek Naming Test (Kosmidis et al., 2012). The Greek Naming Test requires participants to name a series of 40 black and white pictures of common objects, giving a range of scores from 0 to 40 points.

Executive functions

Color Trails Test (CTT; D'Elia et al., 1996). The CTT is a nonalphabetical version of the Trail Making Test (Reitan, 1955) that was developed for cross-cultural populations. In CTT 1, participants are required to connect numbered circles in ascending order. In CTT 2, participants are required to switch between pink and yellow colors while connecting numbers in an ascending order (i.e., pink 1, yellow 2, pink 3, yellow 4, and so on). We recorded the time for the completion for both CTT 1 and 2 in seconds. We also recorded number errors, color errors, self-corrected errors, prompts, and the interference index (ratio of CTT 2 – 1/CTT 1 time scores). English language normative data used to calculate standard performances were derived using US data (D'Elia et al., 1996).

Semantic Verbal Fluency (Kosmidis et al., 2004). Participants were required to generate as many different animal names and 'things you can buy in a supermarket' within 1 min. English language normative data used to calculate standard performances were derived using Australian data (Wardill et al., 2009).

Victoria Stoop Test (Regard, 1981): The Victoria Stoop Test consists of three A4 cards, each containing six rows of four items. In Part D (Dots), participants must name as quickly as possible the color of 24 dots printed in blue, green, red, or yellow. In Part W (Words) the dots are replaced by common words (when, hard, and over), where participants are required to name the colors of the words printed. In Part C (Colors) stimuli are the color names (blue, green, red, and yellow) printed in a color that does not correspond to the name of the color (e.g., "red" is written in blue ink). For each condition, the scores are the time of completion and the number of uncorrected errors. Also, an Interference Index was calculated (Colors/Dot time scores). English language normative data used to calculate standard performances were derived using Canadian data (Troyer et al., 2006).

Procedure

The study was approved by the Monash University Human Research Ethics Committee and was conducted in accordance with the Declaration of Helsinki. All participants volunteered and did not receive financial compensation. Participants were assessed individually in a quiet, distraction-free room in their homes. Home visits were offered as a means of maximizing recruitment potential, participant comfort, and reducing the cost and burden of travel. The first author contacted nine Greek social clubs throughout the Melbourne metropolitan area and was subsequently invited to present information to attendees relating to the normative studies (Staios et al., 2023a, b). Following a general introduction, individuals who showed interest in participation were provided with a plain language statement and allowed to ask additional questions. Those interested in participating provided contact details and a time was arranged to complete the assessment battery. AD participants and their next of kin were informed about the present study by their treating physician(s) during routine clinic attendance and were provided with a plain language statement. Individuals interested in taking part in the study consented to being contacted by the research team to complete the assessment battery. All participants provided informed consent. In instances where individuals with AD were unable to provide informed consent, their next of kin provided consent of their behalf.

Prior to completing cognitive assessment measures, all participants were initially screened using the Greek language adaptions of the MMSE, GDS-15, and GAI, and underwent a semi-structured clinical interview and provided information regarding demographics (e.g., date of birth, education level, years in Australia), medical history (e.g., history of head injury with loss of consciousness, psychiatric or neurological diagnoses), and cultural factors (e.g., age of immigration, English/ Greek and/or other language skills). Healthy participants also provided information regarding self-reported changes in activities of daily living and memory.

Table 1. Demographic data for healthy and AD group

Characteristics	Healthy (*n* = 90) M (SD; range)	AD (*n* = 20) M (SD; range)	*p*	*d*
Age (years)	77.14 (4.46; 70–85)	77.95 (3.50; 72–84)	0.45	−0.18
Sex (male/female)	39/51	9/11	0.89	0.13
Education in Greece (years)	5.60 (0.67; 4–6)	5.70 (0.65; 4–6)	0.55	−0.14
Time in Australia (years)	55.68 (3.94; 46–68)	55.15 (0.98; 54–57)	0.54	0.15
MMSE	26.46 (2.40; 22–30)	16.50 (1.27; 14–20)	< .001	4.82
GDS	1.08 (1.34; 0–5)	1.80 (1.19; 0–4)	0.03	−0.53
GAI	1.68 (1.64; 0–5)	2.65 (1.42; 0–5)	0.01	−0.59
CDR				
Healthy (n)	90	0		
Mild (n)	0	0		
Moderate (n)	0	20		

AD = Alzheimer's disease, CDR = Clinical Dementia Rating Scale, GDS = Geriatric Depression Scale, GAI = Geriatric Anxiety Inventory, MMSE = Mini-Mental State Exam, n = number, SD = standard deviation.

The data presented were collected between May 2018 to February 2020. All tests were administered in Greek by two registered bilingual Greek-English-speaking clinical neuropsychologists. Testing occurred over two sessions, lasting approximately 2 hr in total for healthy participants. Healthy participants completed the test battery reported in the materials section. AD participants completed a shortened test battery, consisting of the following tests: WAIS-IV GR Selected Subtests: Block Design, Matrix Reasoning, Similarities, Digit Span; HVLT-R; WMS-IV Selected Subtests: Visual Reproduction; Greek Story Memory Test; Greek Naming Test; Semantic Verbal Fluency: Animals and Supermarket, and the CTT. Testing occurred over one session and lasted approximately 1 hr for AD participants.

Data analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS), Version 22.0. Independent sample *t* tests were used to examine group differences between healthy and AD participants on age, years of education, time in Australia, MMSE, CDR, GDS-15, and GAI scores. The chi-square test of independence was used to examine group differences regarding the distribution of sex. Cohen's *d* was used to calculate effect sizes with *d* values = 0–0.2 small, 0.2–0.5 medium, and >0.8 large (Cohen, 1988).

Comparison of standardized performances and impairment rates using Greek Australian versus English language normative data

Comparison of standardized performances and impairment rates yielded from newly developed Greek Australian normative data and English language normative data was examined only within healthy participants (*N* = 90). All test raw scores were converted to age adjusted (and where available education adjusted) scaled scores. A series of paired samples *t* tests were used to compare differences between normative data sets on neuropsychological measures within healthy participants. An alpha level of less than .05 two-tailed was considered statistically significant. Cohen's *d* was again used to calculate effects sizes.

The number of participants classified as impaired versus intact were calculated separately using Greek Australian normative data and English language normative data. For both normative data sets, a performance of 1.5 standard deviations below the mean was used to indicate impairment. The relative proportion of the Greek Australian sample scoring 1.5 standard deviations below the mean

was compared to the theoretically expected proportion of 7% of the sample that should fall within the impaired range. A comparison of the proportion of impairment rates when using Greek Australian versus English language normative data were examined using McNemar's tests (McNemar, 1947).

Sensitivity and specificity of AD-related cognitive impairment detection within older Greek Australians

A series of independent sample *t* tests were first conducted to compare raw score performances between healthy and AD participants on all neuropsychological measures. An alpha level of less than .05 two-tailed was considered statistically significant. Cohen's *d* was again used to calculate effect sizes for all neuropsychological measures. A receiver operating characteristic (ROC) curve analysis was used to identify optimal cutoff scores and examine the sensitivity and specificity of the neuropsychological tests using AD participants as the reference standard. Sensitivity and specificity were estimated by the area under the curve (AUC), with a 95% confidence interval. Sensitivity and specificity were calculated for the overall sample with AUC values = < 0.5–0.6 poor, = 0.7–0.8 acceptable, = 0.8–0.9 excellent, and = > 0.9 outstanding (Hosmer et al., 2013). Cut scores for each measure were calculated in order to correspond to the optimal balance between sensitivity and specificity. Sensitivity and specificity of 5th and 10th percentile cut scores according to age adjusted normative data for each measure, were also calculated.

Results

Participants

One hundred and one healthy participants were initially recruited. Five participants were excluded from the study; three scored below cutoff for cognitive impairment on the Greek version of the MMSE; two scored above the cutoff for depression on the GDS-15. A further six participants withdrew due to time constraints as a result of caregiving duties or upcoming medical procedures. The final healthy sample consisted of 90 community-dwelling individuals. A total of 24 participants with a diagnosis of AD were invited to take part in the study. Four participants withdrew and did not complete testing due to a decline in physical and/or mental health. The final sample consisted of 20 participants with a confirmed diagnosis of AD.

Demographic data for the healthy and AD participants are presented in Table 1. In summary, AD and healthy participants did

Table 2. Comparison of scaled scores calculated for healthy participants using Greek Australian versus majority English language norms (N = 90)

Neuropsychological test	Greek Australian norms M (SD; range)	IQR			English language norms M (SD; range)	IQR			p	d
		25th	50th	75th		25th	50th	75th		
WAIS-IV Subtests										
Similarities	10.01 (3.13; 2–18)	8	10	12	4.06 (2.07; 1–9)	3	4	6	< .001	3.91
Vocabulary	9.98 (2.99; 2–18)	8	10	12	4.96 (1.12; 3–8)	4	5	6	< .001	2.43
Information	9.53 (2.11; 5–16)	8	9	11	5.72 (1.87; 2–10)	4	5	7	< .001	4.73
Block Design	10.00 (3.09; 2–18)	8	9	12	8.14 (2.61; 3–13)	7	8	10	< .001	1.43
Matrix Reasoning	9.87 (2.96; 3–18)	8	9.5	12	6.67 (2.09; 1–14)	5	7	8	< .001	2.16
Visual Puzzles	10.15 (3.06; 4–18)	8	9	12	7.68 (1.54; 5–14)	7	7	8	< .001	1.30
Digit Span Total	9.96 (3.01; 2–18)	8	10	12	7.24 (2.07; 3–13)	6	8	8.75	< .001	1.74
Digit Span Forward	9.81 (2.99; 5–18)	8	9	11	6.42 (1.48; 3–12)	6	6	7	< .001	0.73
Digit Span Backward	10.05 (2.91; 2–17)	9	10	12	7.56 (2.02; 2–14)	6	8	9	< .001	2.04
Digit Span Sequencing	9.97 (3.01; 3–18)	9	10	12	8.74 (2.56; 1–15)	7	9	10	< .001	0.84
Arithmetic	9.64 (2.53; 4–16)	8	9	11	7.42 (1.64; 2–11)	6	8	8	< .001	1.69
Symbol Search	9.98 (2.91; 2–17)	8	9.5	12	8.13 (1.71; 3–13)	7	8	9	< .001	0.89
Digit Symbol Coding	9.96 (3.07; 2–18)	8	10	11	6.55 (1.72; 3–12)	6	6	7	< .001	1.91
Hopkins Verbal Learning Test-Revised										
Total Score (T1+T2+T3)	9.73 (2.98; 4–18)	7	9	11	8.67 (1.97; 5–15)	7	8	11	< .001	0.59
Delayed Recall	9.62 (2.87; 2–17)	8	10	12	8.83 (2.21; 1–14)	8	9	11	< .001	0.45
Discrimination Index	9.31 (3.00; 3–15)	7	9	12	9.33 (2.71; 3–14)	8	8	11	0.815	−0.04
WMS-IV Visual Reproduction										
Visual Reproduction I	9.46 (3.04; 3–18)	7	9	12	6.72 (2.91; 1–15)	4	7	9	< .001	3.16
Visual Reproduction II	9.64 (2.89; 5–18)	8	10	11	8.60 (2.46; 1–15)	7	9	10	< .001	0.61
Recognition	9.87 (3.05; 2–18)	8	10	11	7.55 (2.64; 1–15)	6	7	9	< .001	2.13
Verbal Fluency										
Animals	9.61 (3.13; 2–17)	7	9	12	7.74 (1.99; 2–13)	5	5	7	< .001	0.87
Supermarket	10.08 (2.96; 3–18)	8	9	12	7.82 (2.14; 2–15)	6	7	9	< .001	1.03
Victoria Stroop										
Dots	9.22 (2.22; 2–14)	7	10	12	6.30 (3.44; 1–14)	4	7	8	< .001	1.50
Neutral Word	9.41 (2.74; 2–13)	8	10	11	7.01 (2.68; 2–11)	6	7	9	< .001	1.98
Color Word	9.63 (2.47; 3–15)	8	10	11	7.98 (1.74; 4–12)	7	7	10	< .001	0.88
Color Trails Test										
Color Trails Test 1	9.37 (2.90; 2–17)	8	9	11	6.87 (4.33; 2–15)	2	8	10	< .001	0.84
Color Trails Test 2	9.63 (2.47; 2–15)	8	8	12	7.89 (1.74; 2–15)	5	8	10	< .001	0.69

d = effect size, IQR = interquartile range, M = mean, SD = standard deviation.

not differ with respect to age, years of education, sex, or years in Australia. In contrast, there was a significant difference between healthy and AD participants on the MMSE, with the latter group achieving significantly lower scores. All AD participants were classified as moderately impaired on the CDR, while healthy participants fell within normal limits. AD participants endorsed more symptoms of depression than healthy participants, however, no participant in either group reported clinically significant levels of depression or anxiety.

Comparison of standardized performances and impairment rates calculated using Greek Australian versus English language normative data

Standardized performances of our healthy sample, calculated separately using Greek Australian and English language normative data, are presented in Table 2. In summary, mean performances yielded using English language normative data were significantly lower than those obtained using Greek Australian normative data for all tests, with medium to large effect sizes noted. The magnitude of differences between normative data sets generally ranged between one to two standard deviations below the mean when using English language normative data.

Rates of impairment across neuropsychological tests, calculated separately using Greek Australian and English language normative data, are presented in Figure 1 (WAIS-IV subtests) and Figure 2 (all other neuropsychological measures), corresponding to a 1.5 standard deviation criterion. Impairment rates derived from the Greek Australian normative data using a 1.5 standard deviation

criterion showed that rates of impairment generally fell within the expected 7% range. In contrast, impairment rates for all tests derived using English language normative data (except for the HVLT-R Discrimination Index) were significantly higher and ranged from 11-66% (all $ps < .05$).

Sensitivity and specificity of AD-related cognitive impairment detection within older Greek-Australians

Raw neuropsychological test data for all tests are presented separately for healthy and AD participant groups in Table 3. In summary, mean scores for the AD group were significantly lower compared to the healthy group (all $ps < .05$), with all tests displaying medium to large effect sizes. Results from the ROC curve analyses are presented in Table 4 with AUC ranging from .721 to .999 across tests. More specifically, outstanding sensitivity and specificity (AUC > .90) were noted for Block Design and Digit Span (all conditions), Hopkins Verbal Learning Test-Revised (Total Score, Delayed Recall, Percentage Retained, Discrimination Index), Greek Story Memory Test (all conditions), Visual Reproduction II and Semantic Fluency (all conditions). Excellent sensitivity and specificity (AUC > .80) were noted for Matrix Reasoning, Hopkins Verbal Learning Test-Revised (Learning), Visual Reproduction I, and Visual Reproduction Recognition. In contrast, acceptable sensitivity and specificity (AUC > .70) were noted for the Similarities subtest and CTT (all conditions). Sensitivity ranged from .450 to 1.00, with a mean sensitivity of .956 across all measures. Specificity ranged from .511 to .978, with a mean specificity of .844 across all measures.
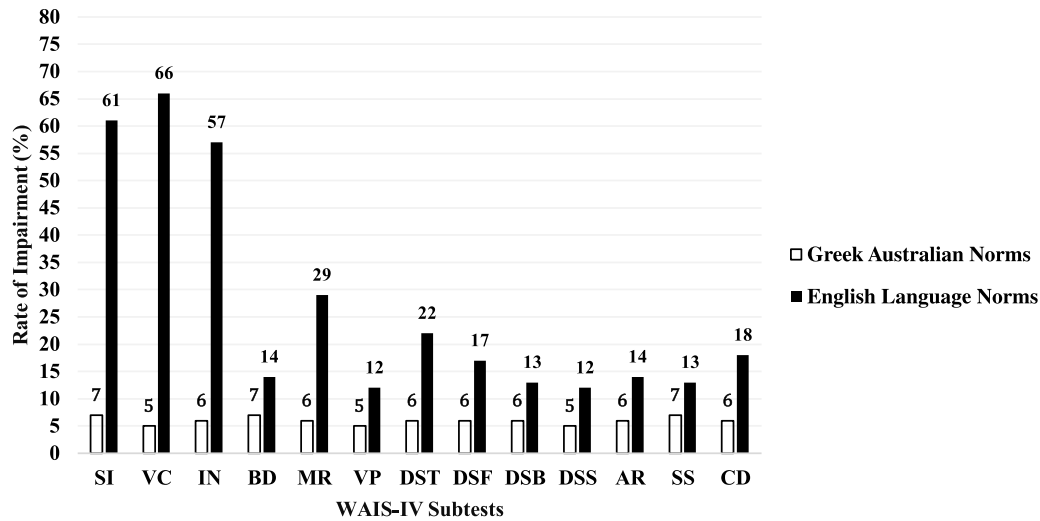
Figure 1. Comparison of rates of impairment on WAIS-IV core subtests calculated using Greek-Australian versus majority English language norms. SI = Similarities, VC = Vocabulary, IN = Information, BD = Block Design, MR = Matrix Reasoning, VP = Visual Puzzles, DS = Digit Span, DSF = Digit Span Forward, DSB = Digit Span Backward, DSS = Digit Span Sequencing, AR = Arithmetic, SS = Symbol Search, CD = Coding.
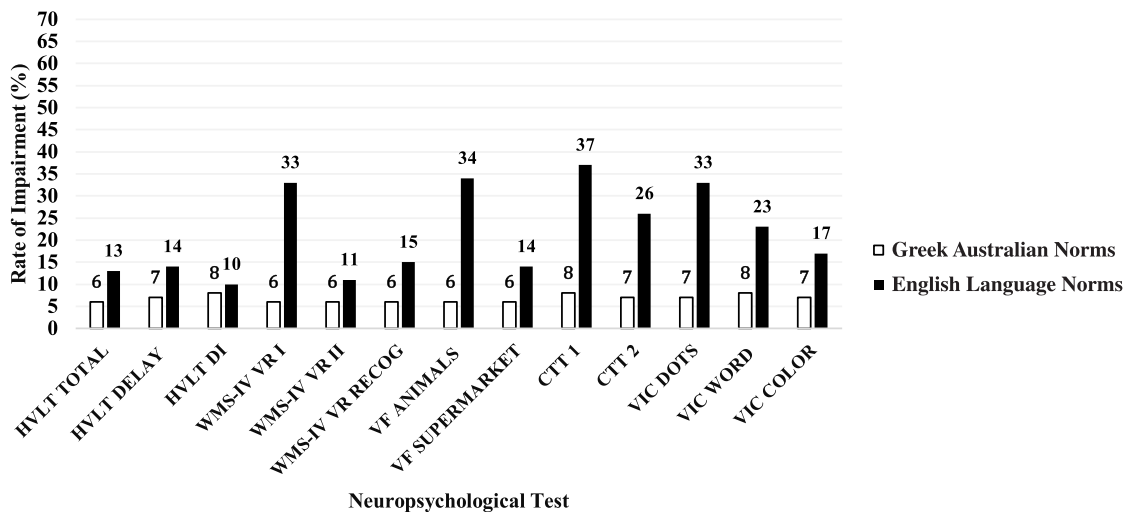

Figure 2. Comparison of rates of impairment across a range of neuropsychological measures calculated using Greek-Australian versus majority English language normative data. HVLT = Hopkins Verbal Learning Test; WMS = Wechsler Memory Scale; VF = Verbal Fluency; CTT = Color Trails Test; VIC = Victoria Stroop.

Sensitivity and specificity of 5th percentile and 10th percentile cut scores are also presented in Table 4. Across measures, sensitivity ranged from .200 to 1.00, with a mean sensitivity of .692. Specificity ranged from .856 to 1.00, with a mean specificity of .926.

Discussion

This study aimed to compare standard performances and impairment rates of a healthy Greek Australian older adult sample calculated using either Greek Australian or English language normative data. We also sought to examine whether cut scores could be identified, capable of sensitively and specifically distinguishing between healthy Greek Australians from those with a diagnosis of AD. Consistent with previous research in other ethnic minority groups (Heaton et al., 2003), the application of English language normative data within our Greek Australian sample resulted in significantly lower performances across all tests

and higher impairment, with medium to large effect sizes noted across most measures. In contrast, when Greek Australian normative data were used, performances yielded typical impairments rates and distributions reflecting expected interindividual variability within a healthy population (Schretlen et al., 2003), with approximately 7% of the sample falling below the 1.5 standard deviation cutoff. While statistical comparisons of education between standardization samples were not possible, the level of education observed within the English language normative samples were substantially higher when compared to the Greek Australian sample; a factor that most likely contributed to the magnitude of the current findings. Similar findings have been noted in previous research, where the application of English language normative data to culturally diverse individuals with ≤6 years of education has been noted to result in high misclassification rates (Cherner et al., 2007). Consistent with previous research, the present findings demonstrate that the use of representative

Table 3. Raw neuropsychological test scores between healthy and AD participants

Neuropsychological test	Healthy (*n* = 90) M (SD)	AD (*n* = 20) M (SD)	*p*	*d*
WAIS-IV Subtests				
Block Design	21.97 (7.44)	7.70 (3.19)	< .001	0.78
Similarities	8.80 (5.20)	3.85 (3.28)	< .001	0.49
Matrix Reasoning	6.39 (2.63)	2.85 (1.95)	< .001	0.60
Digit Span Total Score	18.06 (3.69)	8.90 (2.67)	< .001	0.82
Digit Span Forward	6.59 (1.17)	4.35 (1.13)	< .001	0.69
Digit Span Backward	5.69 (1.48)	2.10 (1.16)	< .001	0.80
Digit Span Sequencing	5.78 (2.11)	2.45 (1.05)	< .001	0.70
Hopkins Verbal Learning Test-Revised				
Total Score (T1+T2+T3)	17.84 (3.61)	9.55 (2.06)	< .001	0.81
Learning	2.95 (1.24)	1.45 (.82)	< .001	0.58
Delayed Recall	5.94 (1.80)	.60 (.75)	< .001	0.89
Percentage Retained	80.37 (18.60)	15.00 (19.96)	< .001	0.86
Discrimination Index	9.41 (1.71)	3.20 (1.19)	< .001	0.90
Greek Story Memory Test				
Immediate Recall	17.02 (5.04)	7.50 (2.16)	< .001	0.77
Delayed Recall	8.66 (2.91)	.55 (.89)	< .001	0.88
Recognition	8.51 (1.41)	3.60 (.88)	< .001	0.90
WMS-IV Visual Reproduction				
Visual Reproduction I	20.41 (6.73)	11.05 (4.85)	< .001	0.62
Visual Reproduction II	12.17 (6.81)	.20 (.89)	< .001	0.77
Recognition	2.80 (1.36)	1.25 (.85)	< .001	0.56
Greek Naming Test				
Total Score	39.62 (.82)	33.90 (2.53)	< .001	0.84
Verbal Fluency				
Animals	12.92 (3.13)	6.70 (1.49)	< .001	0.78
Supermarket	16.77 (3.74)	8.25 (2.24)	< .001	0.81
Color Trails Test				
Color Trails Test 1	92.73 (35.58)	119.75 (40.25)	0.003	0.33
Errors	0.08 (0.46)	2.05 (0.99)	< .001	0.78
Color Trails Test 2	169.68 (61.65)	219.10 (47.40)	0.002	0.41
Color Errors	0.25 (0.48)	3.0 (1.05)	< .001	0.85
Numbers Errors	0.16 (0.45)	2.30 (0.97)	< .001	0.81

d = effect size, IQR = interquartile range, M = mean, SD = standard deviation.

normative data resulted in superior performance of the cognitive measures, with the ability to accurately classify this group of educationally disadvantaged older adults and reduce rates of misclassification (Díaz-Santos et al., 2021; Marquine et al., 2020; Marquine et al., 2021; Scott et al., 2021; Suarez et al., 2021).

We found that the application of English language normative data resulted in significantly lower mean scores and over 57% of healthy participants were classified as impaired on all WAIS-IV verbal subtests. In light of previous research findings, it was not surprising that the use of nonrepresentative normative data resulted in high impairment rates on verbal subtests given the level of education observed within the present sample (Razani et al., 2007; Shuttleworth-Edwards et al., 2004). In contrast, while mean scores on WAIS-IV nonverbal subtests were marginally higher (except for Matrix Reasoning) in comparison to verbal subtests, the use of English language normative data underestimated performances and impairment rates were still unacceptably high. While past research has established that nonverbal measures are also prone to the effects of culture and education, their effects may be less pronounced as such measures are less dependent on language (Ardila et al., 2000). Furthermore, high impairment rates when using English language normative data were also noted on tasks with a timed component, supporting the premise that attitudes toward speeded tasks can be impacted by cultural orientations toward time (Agranovich et al., 2011; Messinis et al., 2011). Interestingly, the application of English language normative data resulted in misclassifying 33% of healthy participants on WMS-IV Visual Reproduction I. This finding is also consistent with previous research examining drawing tests in older low educated immigrant

groups, which have consistently shown that completing such tests is particularly difficult in these populations due to limited experience with using pencils, remembering graphic symbols, and the organization and analysis of visuospatial information (Hong et al., 2011; Nielsen & Jørgensen, 2013; Staios et al., 2022). Finally, no differences were observed on the HVLT-R Discrimination Index between normative data sets. This finding suggests that while cultural and education factors appeared to significantly impact learning and free recall, these factors may have a relatively lower impact on the retention of information and ability to discriminate between true and false positives items. However, further research is needed to confirm these outcomes.

Overall, we were able to establish high sensitivity and specificity in our sample, while also providing cut scores for a broad range of neuropsychological measures. Results showed that Greek Australian participants with AD performed poorer than their healthy counterparts on measures of verbal and visual memory, language, visuospatial skills, and executive functions, with most neuropsychological measures providing robust sensitivity and specificity. Among the neuropsychological tests used in this study, Block Design, Digit Span (total and backward), HVLT-R, Greek Story Memory Test, Greek Naming Test, and Semantic Verbal Fluency were found to have the highest sensitivity and specificity (AUC > .90), followed by Visual Reproduction (AUC > .80). Consistent with previous research (Sala et al., 2017; Salimi et al., 2018; Weissberger et al., 2017), higher sensitivity and specificity was found on tasks assessing verbal and visual memory in discriminating AD, given the profound deficits that those diagnosed with AD exhibit in these cognitive domains (Clarens

Table 4. Sensitivity and specificity of neuropsychological measures and cut points based on ROC curve analysis

Neuropsychological test	AUC (95% CI)	Optimal cutoff			10th %tile			5th %tile		
		Cutoff	SN	SP	Cutoff	SN	SP	Cutoff	SN	SP
WAIS-IV Subtests										
Block Design	.963 (.932–.994)	<13	1.00	0.86	<13	1.00	0.86	<9	0.70	0.96
Similarities	.774 (.678–.870)	<7	0.80	0.69	<2	0.20	0.88	<1	0.20	0.92
Matrix Reasoning	.867 (.780–.954)	<5	0.80	0.73	<3	0.45	0.96	<2	0.25	0.99
Digit Span Total Score	.984 (.965–1.00)	<13	0.95	0.93	<14	0.95	0.89	<13	0.95	0.93
Digit Span Forward	.906 (.833–.979)	<6	0.75	0.92	<5	0.70	0.97	<5	0.70	0.97
Digit Span Backward	.975 (.948–1.00)	<4	0.90	0.97	<4	0.90	0.97	<4	0.90	0.97
Digit Span Sequencing	.912 (.859–.965)	<4	0.80	0.88	<3	0.55	0.90	<2	0.20	0.99
Hopkins Verbal Learning Test-Revised										
Total Score (T1+T2+T3)	.987 (.970–1.00)	<13	0.90	0.96	<13	0.90	0.96	<13	0.90	0.96
Learning	.843 (.749–.937)	<3	0.85	0.61	<1	0.55	1.00	<1	0.50	0.99
Delayed Recall	.995 (.986–1.00)	<3	1.00	0.97	<4	1.00	0.90	<3	1.00	0.97
Percentage Retained	.983 (.962–1.00)	<54	0.95	0.93	<59	0.95	0.91	<47	0.90	0.96
Discrimination Index	.999 (.997–1.00)	<6	1.00	0.99	<7	1.00	0.96	<7	1.00	0.96
Greek Story Memory Test										
Immediate Recall	.944 (.902–.986)	<11	0.90	0.90	<11	0.90	0.90	<9	0.75	0.96
Delayed Recall	.996 (.989–1.00)	<4	1.00	0.92	<5	1.00	0.89	<4	1.00	0.92
Recognition	.998 (.995–1.00)	<6	1.00	0.98	<7	1.00	0.91	<7	1.00	0.91
WMS-IV Visual Reproduction										
Visual Reproduction I	.874 (.791–.957)	<15	0.80	0.84	<12	0.70	0.93	<11	0.50	0.94
Visual Reproduction II	.985 (.965–1.00)	<4	0.95	0.93	<4	0.95	0.93	<4	0.95	0.93
Recognition	.829 (.746–.912)	<3	0.95	0.59	<13	1.00	0.86	<9	0.70	0.96
Greek Naming Test										
Total Score	.968 (.906–1.00)	<38	0.95	0.96	<38	0.95	0.96	<38	0.95	0.96
Verbal Fluency										
Animals	.973 (.948–.998)	<10	1.00	0.86	<9	0.85	0.92	<9	0.85	0.92
Supermarket	.978 (.951–1.00)	<12	0.95	0.94	<12	0.95	0.94	<12	0.95	0.94
Color Trails Test										
Color Trails Test 1	.729 (.616–.842)	>81	0.85	0.51	>137	0.45	0.90	>153	0.25	0.96
Errors	.960 (.900–1.00)	>0	0.95	0.94	0	0.00	0.00	>0	0.95	0.94
Color Trails Test 2	.734 (.630–.838)	>168	0.90	0.54	>257	0.40	0.91	>299	0.00	0.94
Color Errors	.996 (.987–1.00)	>1	1.00	0.98	>0	1.00	0.77	>0	1.00	0.77
Numbers Errors	.956 (.895–1.00)	>1	0.85	0.97	>0	0.95	0.87	>0	0.95	0.87

AUC = area under the curve, ROC = receiver operating characteristic, SN = sensitivity, SP = specificity.

et al., 2022; Ramirez-Gomez et al., 2017). Furthermore, measures of executive functioning, namely working memory and verbal fluency, showed equivalent sensitivity and specificity. Previous research examining dementia profiles in ethnic minority groups have noted that verbal memory deficits outweigh those observed in visuospatial skills, language and executive functioning, in the initial stage of disease progression (Barnes-Marrero et al., 2022; Nielsen et al., 2018). In this context, the degree of executive deficits observed in our AD group is likely attributed to the fact that they were moderately impaired, where such deficits are commonly observed (Weintraub et al., 2012).

Moreover, of all the neuropsychological measures administered, the Similarities subtest and the CTT showed low specificity (AUC = <.80). Poor specificity values noted on the Similarities subtest were likely due to factors such as level and quality of education obtained in Greece pre-immigration and acculturation (Staios, 2022). Regarding the CTT, results revealed that the total time score for both conditions was found to yield a high proportion of false positive results, as indicated by low specificity values, leading to a misidentification of approximately half of the healthy participants. Similar findings have previously been noted in a sample of educationally disadvantaged Brazilian older adults, where both conditions of the CTT time score for were found to lack sensitivity and specificity when compared to AD participants (Araujo et al., 2020). These findings are likely due to several factors, including no previous experience with psychometric testing, level and quality of education, and cultural orientation toward time and speeded tasks (Agranovich et al., 2011; Al-Jawahiri & Nielsen,

2021; Messinis et al., 2011). In light of these findings, in instances where clinicians are required to carry out evaluations of Greek Australian elders with moderate dementia, we recommend using the following measures: *WAIS-IV Block Design and Digit Span, Hopkins Verbal Learning Test-Revised, Greek Story Memory Test, Greek Naming Test, and Verbal Fluency.* The aforementioned measures displayed excellent sensitivity and specificity in discriminating health from AD participants and are also appropriate for testing several cognitive domains known to be impacted by AD (Lezak et al., 2012).

Finally, we aimed to identify cut scores for a broad range of neuropsychological measures, based on the optimal balance between sensitivity and specificity, and compare this to traditionally used benchmarks of 5th and 10th percentiles. Overall, optimal cut scores yielded excellent results for the assessment of moderately impaired individuals with AD and are now available for research and clinical use. Inspection of the data concerning the 5th percentile revealed that this cutoff compromised sensitivity across most tasks. In contrast, the cut scores across both the suggested optimal and 10th percentile ranges remained approximately the same. However, a notable difference was observed for the Similarities and Matrix Reasoning subtests, Visual Reproduction (recognition), and CTT, where the use of 5th and 10th percentiles corresponded to a higher discrepancy between sensitivity and specificity in favor of the latter in comparison with the optimal cutoff scores. This finding may provide useful information in clinical practice, as it allows the selection of the cutoff score with the most suitable combination between sensitivity and specificity

depending on the clinical question at hand. In other words, a high sensitivity score may be preferable when older adults at higher risk for dementia need to be identified, while a high specificity score could be more useful when confirmation of AD is necessary. In summary, the selection of the cutoff score for these specific tests could provide different outcomes according to the goal of the clinical assessment.

This study had several limitations. First, participants were selected through convenience sampling and consisted of primary school educated individuals. Therefore, results are not generalizable or appropriate for clinical use when assessing Greek Australians with higher levels of education. Second, the fact that our clinical group comprised moderately impaired AD participants, as can be inferred by the discrepancy in MMSE and CDR scores between the two groups, may partially explain the excellent sensitivity and specificity observed in several neuropsychological tests. Therefore, the cut scores derived in this study should only be applied to Greek Australians presenting with moderate to severe forms of dementia. In cases where milder forms of dementia are suspected, we advise referring to our normative studies and using these data to calculate probable levels of impairment (Staios et al., 2023a, b). Despite our best efforts to recruit participants with mild cognitive impairment, we were unable to do so. Greek Australians and other culturally diverse groups continue to be underrepresented in clinical research, relative to their Anglo Australian counterparts (Low et al., 2019). Factors such as stigma, limited access to culturally appropriate resources and treatment services continue to be an issue facing culturally diverse Australians, leading to underutilization of health care services (LoGiudice et al., 2001; Low et al., 2010; Phillipson et al., 2015). As a result, a combination of these factors impacted our ability to recruit participants with early-stage AD. Future research would benefit from exploring the sensitivity and specificity of the tests used in the present study within a broader and milder range of clinical presentations, as well as their ability to differentiate between a range of neurocognitive disorders within the Greek Australian community.

In conclusion, we believe our findings represent an important contribution to the field of clinical neuropsychology. We have demonstrated how using English language normative data within healthy educationally disadvantaged ethnic minority groups can result in erroneous diagnostic outcomes. We have also demonstrated how use of demographically focused normative data and alternate cut points can help ameliorate this issue. We anticipate that the methodology employed in the present study may be used as a template for other ethnic minority groups to improve cross-cultural neuropsychological practice and test development internationally. Finally, while culturally specific tests and normative data may facilitate accuracy of testing outcomes, they should not be viewed as the only solution to improving the validity of cognitive assessment within underrepresented groups (Ardila, 2005). The need for increasing cultural competence, the development of culturally considered clinical guidelines, ongoing professional development, and wider representation within the field of neuropsychology are necessary (Rivera Mindt et al., 2010).

References

Agranovich, A. V., Panter, A. T., Puente, A. E., & Touradji, P. (2011). The culture of time in neuropsychological assessment exploring the effects of culture-specific time attitudes on timed test performance in Russian and American samples. *Journal of the International Neuropsychological Society*, *17*(4), 692–701. https://doi.org/10.1017/S1355617711000592

Al-Jawahiri, F., & Nielsen, T. R. (2021). Effects of acculturation on the cross-cultural neuropsychological test battery (CNTB) in a culturally and linguistically diverse population in Denmark. *Archives of Clinical Neuropsychology*, *36*(3), 381–393. https://doi.org/10.1093/arclin/acz083

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders.* (5th edn). American Psychiatric Association.

Araujo, N. B., Nielsen, T. R., Barca, M. L., Engedal, K., Marinho, V., Deslandes, A. C., Coutinho, E. S., & Laks, J. (2020). Brazilian version of the European Cross-Cultural Neuropsychological Test Battery (CNTB-BR): diagnostic accuracy across schooling levels. *Revista Brasileira de Psiquiatria*, *42*(3), 286–294. https://doi.org/10.1590/1516-4446-2019-0539

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review*, *15*(4), 185–195. https://doi.org/10.1007/s11065-005-9180-y

Ardila, A., Ostrosky-Solis, F., Rosselli, M., & Gómez, C. (2000). Age-related cognitive decline during normal aging: The complex effect of education. *Archives of Clinical Neuropsychology*, *15*(6), 495–513.

Australian Bureau of Statistic (2016). *People in Australia who were born in Greece.* Canberra: Author.

Australian Institute of Health and Welfare (2018). *Dementia in Australia.* Canberra: Author.

Australian Institute of Health and Welfare (2021). *Older Australians.* Canberra: Author.

Barnes-Marrero, I., Horter, L., Hayden, J. D., Patel, N. C., Mendoza, L., & Castillo, L. (2022). Diagnostic accuracy of the repeatable battery of the assessment of neuropsychological status update, Spanish version, in predicting Alzheimer's disease among Hispanic older adults in the United States reporting memory problems. *Applied Neuropsychology: Adult*, *29*(4), 509–519. https://doi.org/10.1080/23279095.2020.1777554

Blakemore, A., Kenning, C., Mirza, N., Daker-White, G., Panagioti, M., & Waheed, W. (2018). Dementia in UK South Asians: A scoping review of the literature. *BMJ Open*, *8*(4), e020290. https://doi.org/10.1136/bmjopen-2017-020290

Brandt, J., & Benedict, R. H. B. (2001). *Hopkins verbal learning test – revised. Administration manual.* Psychological Assessment Resources.

Cherner, M., Suarez, P., Lazzaretto, D., Fortuny, L. A., Mindt, M. R., Dawes, S., Marcotte, T., Grant, I., Heaton, R., & HNRC group (2007). Demographically corrected norms for the brief visuospatial memory test-revised and hopkins verbal learning test-revised in monolingual Spanish speakers from the U.S.-Mexico border region. *Archives of Clinical Neuropsychology*, *22*(3), 343–353. https://doi.org/10.1016/j.acn.2007.01.009

Clarens, M. F., Crivelli, L., Calandri, I., Chrem Méndez, P., Martin, M. E., Russo, M. J., & Allegri, R. F. (2022). Neuropsychological profile of Alzheimer's disease based on amyloid biomarker findings results from a South American cohort. *Applied Neuropsychology: Adult*, *29*(3), 345–350. https://doi.org/10.1080/23279095.2020.1756816

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillside, NJ: Lawrence Erlbaum Associates.

D'Elia, L. F., Satz, P., Uchiyama, C. L., & White, T. (1996). *Color Trails Test.* Odessa, Florida: Psychological Assessment Resources.

Daugherty, J. C., Puente, A. E., Fasfous, A. F., Hidalgo-Ruzzante, N., & Pérez-Garcia, M. (2017). Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests. *Applied Neuropsychology: Adult*, *24*(1), 16–22. https://doi.org/10.1080/23279095.2015.1036992

Díaz-Santos, M., Suárez, P. A., Marquine, M. J., Umlauf, A., Rivera Mindt, M., Fortuny, Artiola I., Heaton, L., R., K., & Cherner, M. (2021). Updated demographically adjusted norms for the brief visuospatial memory test-revised and hopkins verbal learning test-revised in Spanish-speakers from the U.S.-Mexico border region: The NP-NUMBRS project. *The Clinical Neuropsychologist*, *35*(2), 374–395. https://doi.org/10.1080/13854046.2020.1861329

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., *et al.* (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *The Lancet Neurology*, 6(8), 734–746. https://doi.org/10.1016/s1474-4422(07)70178-3

Fanany, R., & Avgoulas, M. I. (2019). Greek Identity in Australia. In S. Ratuva (Eds.), *The Palgrave Handbook of Ethnicity* (pp. 1186–1199). Palgrave Macmillan.

Fountoulakis, K. N., Tsolaki, M., Chantzi, H., & Kazis, A. (2000). Mini mental state examination (MMSE): A validation study in Greece. *American Journal of Alzheimer's Disease*, 15(6), 342–345. https://doi.org/10.1177/153331750001500604

Fountoulakis, K. N., Tsolaki, M., Iacovides, A., Yesavage, J., O'Hara, R., Kazis, A., *et al.* (1999). The validation of the short form of the geriatric depression scale (GDS) in Greece. *Aging: Clinical and Experimental Research*, 11(6), 367–372. https://doi.org/10.1007/bf03339814

Franzen, S., Papma, J. M., van den Berg, E., & Nielsen, T. R. (2021). Cross-cultural neuropsychological assessment in the European union: A Delphi expert study. *Archives of Clinical Neuropsychology*, 36(5), 815–830. https://doi.org/10.1093/arclin/acaa083

Fratti, S., Bowden, S.C., & Pino, O. (2011). Diagnostic memory assessment in Italian-Born Australians. International Psychogeriatrics, 23:7,1133–43.

Heaton, R. K., Taylor, M. J., & Manly, J. (2003). Demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III. In D. S. Tulsky, D. H. Saklofske, G. J. Chelune, R. K. Heaton, R. J. Ivnik, R. Bornstein, A. Prifitera, & M. F. Ledbetter (Eds.), *Clinical interpretation of the WAIS-III and WMS-III* (pp. 181–210). Academic Press.

Hester, R. L., Kinsella, G. J., Ong, B., & Turner, M. (2004). Hopkins verbal learning test: Normative data for older Australian adults. *Australian Psychologist*, 39(3), 251–255. https://doi.org/10.1080/000500604123312950

Hong, Y. J., Yoon, B., Shim, Y. S., Cho, A.-H., Lee, E.-S., Kim, Y.-I., & Yang, D. W. (2011). Effect of literacy and education on the visuoconstructional ability of non-demented elderly individuals. *Journal of the International Neuropsychological Society*, 17(05), 934–939. https://doi.org/10.1017/s1355617711000889

Hosmer, W. D., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression (3rd ed)*. New Jersey: Wiley, https://doi.org/10.1002/9781118548387

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*, 140(6), 566–572. https://doi.org/10.1192/bjp.140.6.566

Jupp, J. (2001). *The Australian people: An encyclopedia of the nation, and their origins*. New York: Cambridge University Press.

Kochhann, R., Varela, J. S., de Macedo Lisboa, C. S., & Chaves, M. L. F. (2010). The mini mental state examination: Review of cutoff points adjusted to schooling in a large southern Brazilian sample. *Dementia & Neuropsychologia*, 4(1), 35–41. https://doi.org/10.1590/S1980-57642010DN40100006

Kosmidis, M. H. (2017). Challenges in the neuropsychological assessment of illiterate older adults. *Language, Cognition and Neuroscience*, 33(3), 373–386. https://doi.org/10.1080/23273798.2017.1379

Kosmidis, M. H., Bozikas, V. P., & Vlahou, C. H. (2012). *Neuropsychological Test Battery*. Greece: Lab of Cognitive Neuroscience, School of Psychology, Aristotle University of Thessaloniki.

Kosmidis, M. H., Vlahou, C. H., Panagiotaki, P., & Kiosseoglou, G. (2004). The verbal fluency task in the Greek population: Normative data, and clustering and switching strategies. *Journal of the International Neuropsychological Society*, 10(2), 164–172. https://doi.org/10.1017/S1355617704102014

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th edn). New York: Oxford University Press.

LoGiudice, D., Hassett, A., Cook, R., Flicker, L., & Ames, D. (2001). Equity of access to a memory clinic in Melbourne? Non-English speaking background attenders are more severely demented and have increased rates of psychiatric disorders. *International Journal of Geriatric Psychiatry*, 16(3), 327–334. https://doi.org/10.1002/gps.346

Low, L. F., Anstey, K. J., Lackersteen, S. M., Camit, M., Harrison, F., Draper, B., & Brodaty, H. (2010). Recognition, attitudes and Causal beliefs regarding dementia in italian, Greek and Chinese Australians. *Dementia and Geriatric Cognitive Disorders*, 30(6), 499–508. https://doi.org/10.1159/000321667

Low, L. F., Barcenilla-Wong, A. L., & Brijnath, B. (2019). Including ethnic and cultural diversity in dementia research. *Medical Journal of Australia*, 211(8), 345. https://doi.org/10.5694/mja2.50353

Marquine, M. J., Morlett Paredes, A., Madriaga, C., Blumstein, Y., Umlauf, A., Kamalyan, L., Cherner, M., *et al.* (2020). Demographically-adjusted norms for selected tests of verbal fluency: Results from the neuropsychological norms for the US-Mexico border region in Spanish (NP-NUMBRS) project. *The Clinical Neuropsychologist*, 1–24. https://doi.org/10.1080/13854046.2020.1762

Marquine, M. J., Yassai-Gonzalez, D., Perez-Tejada, A., Umlauf, A., Kamalyan, L., Morlett Paredes, A., Rivera Mindt, M., Artiola, I., Fortuny, L., Cherner, M., & Heaton, R. K. (2021). Demographically adjusted normative data for the Wisconsin card sorting test-64 item: Results from the neuropsychological norms for the U.S.-Mexico border region in Spanish (NP-NUMBRS) project. *The Clinical Neuropsychologist*, 35(2), 339–355. https://doi.org/10.1080/13854046.2019.1703042

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153–157. https://doi.org/10.1007/BF02295996

Messinis, L., Malegiannaki, A.-C., Christodoulou, T., Panagiotopoulos, V., & Papathanasopoulos, P. (2011). Color trails test: Normative data and criterion validity for the Greek adult population. *Archives of Clinical Neuropsychology*, 26(4), 322–330. https://doi.org/10.1093/arclin/acr027

Nielsen, T. R., Segers, K., Vanderaspoilden, V., Bekkhus-Wetterberg, P., Minthon, L., Pissiota, A., Bjørkløf, G. H., Beinhoff, U., Tsolaki, M., Gkioka, M., & Waldemar, G. (2018). Performance of middle-aged and elderly European minority and majority populations on a cross-cultural neuropsychological test battery (CNTB). *The Clinical Neuropsychologist*, 32(8), 1411–1430. https://doi.org/10.1080/13854046.2018.1430256

Nielsen, T. R., & Jørgensen, K. (2013). Visuoconstructional abilities in cognitively healthy illiterate Turkish immigrants: A quantitative and qualitative investigation. *The Clinical Neuropsychologist*, 27(4), 681–692. https://doi.org/10.1080/13854046.2013.767379

Nielsen, T. R., & Waldemar, G. (2021). Cross-cultural cognitive examination in aging migrants. In M. E. Alaoui-Faris, A. Federico, & W. Grisold (Eds.), *Neurology in Migrants and Refugees* (1st ed., pp. 267–281). Springer Nature: Switzerland.

Noutsos, H. (2003). *The road of the camel and school: The educational policy in Greece from 1944-46*. Athens: Bibliorama.

Pachana, N. A., Byrne, G. J., Siddle, H., Koloski, N., Harley, E., & Arnold, E. (2007). Development and validation of the geriatric anxiety inventory. *International Psychogeriatrics*, 19(01), 103–114.

Pedraza, O., Clark, J. H., O'Bryant, S. E., Smith, G. E., Ivnik, R. J., Graff-Radford, N. R., Lucas, J. A., Willis, F. B., Petersen, R. C., & Lucas, J. A. (2011). Diagnostic validity of age and educa- tion corrections for the mini-mental state examination in older African Americans. *Journal of the American Geriatrics Society*, 60(2), 328–331. https://doi.org/10.1111/j.1532-5415.2011.03766.x

Phillipson, L., Magee, C., Jones, S. C., Reis, S., & Skladzien, E. (2015). Dementia attitudes and help-seeking intentions: An investigation of responses to two scenarios of an experience of the early signs of dementia. *Aging and Mental Health*, 19(11), 968–977. https://doi.org/10.1080/13607863.2014.995588

Pienaar, I., Shuttleworth-Edwards, A., Klopper, C., & Radloff, S. (2016). Wechsler adult intelligence scale-fourth edition preliminary normative guidelines for educationally disadvantaged xhosa-speaking individuals. *South African Journal of Psychology*, 47(2), 159–170. https://doi.org/10.1177/0081246316654805

Plitas, A., Tucker, A., Kritikos, A., Walters, I., & Bardenhagen, F. (2009). Comparative study of the cognitive performance of Greek Australian and Greek national elderly: Implications for neuropsychological practice. *Australian Psychologist*, 44(1), 27–39. https://doi.org/10.1080/00050060802587694

Ramirez-Gomez, L., Zheng, L., Reed, B., Kramer, J., Mungas, D., Zarow, C., & Chui, H. (2017). Neuropsychological profiles differentiate Alzheimer disease from subcortical ischemic vascular dementia in an autopsy-defined cohort. Dementia and geriatric cognitive. *Dementia and Geriatric Cognitive Disorders*, 44(1-2), 1–11. https://doi.org/10.1159/000477344

Razani, J., Murcia, G., Tabares, J., & Wong, J. (2007). The effects of culture on WASI test performance in ethnically diverse individuals. *The*

Clinical Neuropsychologist, *21*(5), 776–788. https://doi.org/10.1080/13854040701437481

Regard, M. (1981). *Cognitive rigidity and flexibility: A neuropsychological study. [Unpublished doctoral dissertation].* British Columbia, Canada: University of Victoria.

Reitan R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology*, *19*(5), 393–394. https://doi.org/10.1037/h0044509

Rivera Mindt, M., Byrd, D., Saez, P., & Manly, J. (2010). Increasing culturally competent neuropsychological services for ethnic minority populations: A call to action. *The Clinical Neuropsychologist*, *24*(3), 429–453. https://doi.org/10.1080/13854040903058960

Rivera Mindt, M., Marquine, M. J., Aghvinian, M., Paredes, A. M., Kamalyan, L., Suárez, P., Heaton, A., Scott, T. M., Gooding, A., Diaz-Santos, M., Umlauf, A., Taylor, M. J., Artiola I.Fortuny, L., Heaton, R. K., & Cherner, M. (2020). The Neuropsychological Norms for the U.S.-Mexico border region in Spanish (NP-NUMBRS) project: Overview and considerations for life span research and evidence-based practice. *The Clinical Neuropsychologist*, 1–15. https://doi.org/10.1080/13854046.2020.1794

Sala, I., Illán-Gala, I., Alcolea, D., Sánchez-Saudinós, M. B., Salgado, S. A., Morenas-Rodríguez, E., Subirana, A., Videla, L., Clarimón, J., Carmona-Iragui, M. D., Ribosa-Nogué, R., Blesa, R., Fortea, J., & Lleó, A. (2017). Diagnostic and prognostic value of the combination of two measures of verbal memory in mild cognitive impairment due to Alzheimer's disease. Journal of Alzheimer's. *Journal of Alzheimer's Disease*, *58*(3), 909–918.

Salimi, S., Irish, M., Foxe, D., Hodges, J. R., Piguet, O., & Burrell, J. R. (2018). Can visuospatial measures improve the diagnosis of Alzheimer's disease? Alzheimer's & dementia: Diagnosis. *Assessment & Disease Monitoring*, *10*(1), 66–74. https://doi.org/10.1016/j.dadm.2017.10.004

Schretlen, D. J., Munro, C. A., Anthony, J. C., & Pearlson, G. D. (2003). Examining the range of normal intraindividual variability in neuropsychological test performance. *Journal of the International Neuropsychological Society*, *9*(06), 864–870. https://doi.org/10.1017/s1355617703960061

Scott, T. M., Morlett Paredes, A., Taylor, M. J., Umlauf, A., Fortuny, Artiola I., Heaton, L., R., K., Cherner, M., Marquine, M. J., & Rivera Mindt, M. (2021). Demographically-adjusted norms for the WAIS-R block design and arithmetic subtests: Results from the neuropsychological norms for the US-Mexico border region in spanish (NP-NUMBRS) project. *The Clinical neuropsychologist*, *35*(2), 419–432. https://doi.org/10.1080/13854046.2019.1707285

Shuttleworth-Edwards, A. B. (2016). Generally representative is representative of none: Commentary on the pitfalls of IQ test standardization in multicultural settings. *The Clinical Neuropsychologist*, *30*(7), 975–998.

Shuttleworth-Edwards, A. B., Kemp, R. D., Rust, A. L., Muirhead, J. G. L., Hartman, N. P., & Radloff, S. E. (2004). Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III test performance. *Journal of Clinical and Experimental Neuropsychology*, *26*(7), 903–920. https://doi.org/10.1080/13803390490510824

Staios, M. (2022). Culturally aware neuropsychological assessment with Greek immigrants. In F. Irani (Eds.), *Cultural diversity in neuropsychological assessment* (pp. 341–353). Routledge Inc.

Staios, M., Kosmidis, M. H., Kokkinis, N., Papadopoulos, A., Nielsen, T. R., Kalinowski, P., March, E., & Stolwyk, R. J. (2023a). The Greek Australian neuropsychological normative study: Neuropsychological tests & norms for Greek Australians aged 70–85 years. *Australian Psychologist*, 1–15. https://doi.org/10.1080/00050067.2022.2151337

Staios, M., Kosmidis, M. H., Nielsen, T. R., Papadopoulos, A., Kokkinis, N., Stogiannidou, A., March, E., & Stolwyk, R. J. (2023b). The wechsler adult intelligence scale-fourth edition, Greek adaptation (WAIS-IV GR): Confirmatory factor analysis and specific reference group normative data for Greek Australian older adults. *Australian Psychologist*, 1–16. https://doi.org/10.1080/00050067.2023.2179387

Staios, M., Nielsen, T. R., Kosmidis, M. H., Papadopoulos, A., Kokkinias, A., Velakoulis, D., Tsiaras, Y., March, E., & Stolwyk, R. J. (2022). Validity of Visuoconstructional Assessment Methods within Healthy Elderly Greek Australians: Quantitative and Error Analysis. *Archives of Clinical Neuropsychology*, *38*(4), 598–607. https://doi.org/10.1093/arclin/acac091

Suarez, P. A., Díaz-Santos, M., Marquine, M. J., Kamalyan, L., Mindt, M. R., Umlauf, A., Heaton, R. K., Grant, I., & Cherner, M. (2021). Demographically adjusted norms for the trail making test in native Spanish speakers: Results from the neuropsychological norms for the US-Mexico border region in Spanish (NP-NUMBRS) project. *The Clinical neuropsychologist*, *35*(2), 308–323. https://doi.org/10.1080/13854046.2020.1800099

Troyer, A. K., Leach, L., & Strauss, E. (2006). Aging and response inhibition: Normative data for the victoria stroop test. *Aging, Neuropsychology, and Cognition*, *13*(1), 20–35. https://doi.org/10.1080/138255890968187

Wallace, K., Berry, J., & Shores, E. A. (2018). The Chinese Australian Neuropsychological Normative Study: Neuropsychological test norms for Chinese Australians aged 55–90 years. *Australian Psychologist*, *53*(5), 427–443. https://doi.org/10.1111/ap.12340

Walterfang, M., Velakoulis, D., Gibbs, A., & Lloyd, J. (2003). The NUCOG: Construction and piloting of a cognitive screening instrument in a neuropsychiatric unit. *Australasian Psychiatry*, *11*(3), 325–329. https://doi.org/10.1046/j.1440-1665.2003.t01-1-00558.x

Wardill, T., Anderson, V., Graham, A., & Perre, C. (2009). *Normative data for older Australians*. Australia: Box Hill, Monash University.

Wechsler, D. (2008). *Wechsler adult intelligence scale-fourth edition*. San Antonio, Texas: Pearson.

Wechsler, D. (2009). *Wechsler memory scale-fourth edition*. San Antonio, Texas: Pearson.

Wechsler, D. (2014). *Wechsler adult intelligence scale-fourth edition, greek adaptation*. Athens: Motibo Publishing.

Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, *2*(4), a006171. https://doi.org/10.1101/cshperspect.a00617

Weissberger, G. H., Strong, J. V., Stefanidis, K. B., Summers, M. J., Bondi, M. W., & Stricker, N. H. (2017). Diagnostic accuracy of memory measures in Alzheimer's dementia and mild cognitive impairment: A systematic review and. *Neuropsychology Review*, *27*(4), 354–388. https://doi.org/10.1007/s11065-017-9360-6

World Health Organisation (2022). *Dementia*, Geneva: Author.