THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

A comparison of the Greek ACE-III, M-ACE, ACE-R, MMSE and ECAS in the assessment and identification of Alzheimer's Disease

OPEN ACCESS

A comparison of the Greek ACE-III, M-ACE, ACE-R, MMSE and ECAS in the assessment and identification of Alzheimer's Disease

Panagiotis Kourtesis[a,b,c]*, Eleni Margioti[d,e], Christina Demenega[e], Foteini Christidi[f] and Sharon Abrahams[a,g].

[a] *Department of Psychology, Human Cognitive Neuroscience, University of Edinburgh, Edinburgh, UK;*

[b] *Lab of Experimental Psychology, Suor Orsola Benincasa University of Naples, Naples, Italy;*

[c] *Interdepartmental Centre for Planning and Research "Scienza Nuova", Suor Orsola Benincasa University of Naples, Naples, Italy;*

[d] *Department of Psychology, Aristotle University of Thessaloniki, Thessaloniki, Greece;*

[e] *Department of Psychology, Athens Alzheimer's Association, Athens, Greece;*

[f] *A' Department of Neurology, National and Kapodistrian University of Athens, Athens, Greece;*

[g] *Euan MacDonald Centre for Motor Neurone Disease Research, Royal Infirmary of Edinburgh, Edinburgh, UK;*

* Panagiotis Kourtesis

7 George Square, Edinburgh, EH8 9JZ,

Scotland, United Kingdom

Email: pkourtes@exseed.ed.ac.uk

Abstract

Objective: This study aimed to adapt the Addenbrooke's Cognitive Examination-III (ACE-III) and Mini-ACE (M-ACE) into Greek and then to examine the convergent validity against their predecessors Addenbrooke's Cognitive Examination-Revised (ACE-R) and Mini-Mental State Examination (MMSE) in a Greek population. Moreover, a primary aim was to appraise the utility of each screen by conducting a comparison of the psychometric properties of ACE-III, M-ACE, ACE-R, MMSE, and the Edinburgh Cognitive and Behavioural ALS Screen (ECAS) in detecting Alzheimer's disease (AD).

Methods: Forty patients with AD were recruited and matched with 38 controls. Bayesian Pearson's correlation analysis was conducted to examine the convergent validity. Receiver operating characteristic curve analysis was implemented to appraise the sensitivity and specificity of the tests in detecting AD.

Results: The ACE-III, M-ACE, and the ECAS scores robustly correlated with ACE-R and MMSE. The ACE-III and the ECAS-ALS Non-Specific score were the most sensitive and specific tools in detecting AD, closely followed by ECAS Total score and M-ACE. Only ECAS Total score correlated with the duration of disease. The ECAS scores were more resilient to ceiling effects than the other screens. M-ACE produced fewer ceiling effects than MMSE.

Conclusion: The Greek ACE-III and M-ACE were successfully adapted and showed good convergent validity against their predecessors. They showed very good psychometric properties in detecting AD and may be considered in hectic clinical settings. ECAS Total score and ECAS-ALS Non-Specific showed comparable psychometric properties in the detection of AD and may be considered in poly-pathological clinics where motor impairments are common.

Keywords: Greek; ACE-III; M-ACE; ECAS; Alzheimer's disease; motor disabilities;

Introduction

Cognitive assessment is crucial for the detection of Alzheimer's disease (AD) and for the differential diagnosis of other types of dementia, such as frontotemporal dementia (FTD) (Fields *et al.*, 2011). Comprehensive assessment of cognition and behaviour has clinical implications for patient care, regarding the available treatment options, survival expectancy, competency to drive or provide informed consent, ability to live independently at home, the carer's burden and quality of life (Hsieh *et al.*, 2015). However, in hectic clinical settings, briefer cognitive screening methods are often the test of choice, with patients with more complex needs or diagnostic uncertainties being referred for full neuro-psychological assessment (Hsieh *et al.*, 2015).

Addenbrooke's Cognitive Examination Revised (ACE-R) and the embedded Mini-Mental State Examination (MMSE) are the predominant brief screening tests for dementia in the Greek population, with administration times of approximately 15 and 5 min, respectively (Konstantinopoulou *et al.*, 2011). Both of them were designed to briefly examine a wide range of cognitive domains: attention, memory, language, visuo-spatial components and verbal fluency (Mioshi *et al.*, 2006). ACE-R aids in the detection, differentiation and monitoring of cognitive decline in dementia syndromes, such as FTD and AD (Hsieh *et al.*, 2012; Kipps *et al.*, 2008; Mathew *et al.*, 2011; Raimondi *et al.*, 2012).

However, ACE-R has several limitations (Hsieh *et al.*, 2013); for example, healthy adults repeatedly fail on the verbal repetition item, which might be a result of hearing problems or distraction (Hsieh *et al.*, 2013; Valcour *et al.*, 2002), and ceiling effects have been observed in the measure of comprehension (Brugnolo *et al.*, 2009). Acknowledging these weaknesses led to the development of Addenbrooke's Cognitive Examination III (ACE-III). While the ACE-III does not incorporate MMSE, it continues to assess the same five cognitive domains, with new items in verbal repetition and language comprehension tasks, while backward spelling was replaced by serial 7s subtraction (Brugnolo et al., 2009; Ganguli et al., 1990; Hsieh et al., 2013; Valcour et al., 2002).

ACE-III has been validated against extensive neuro-psychological tests (Hsieh *et al.*, 2015; Hsieh *et al.*, 2013). However, even ACE-III, which demands 15–20 min to administer, has been suggested to be excessive for some busy clinical settings (Hsieh *et al.*, 2015). Mini-Addenbrooke's Cognitive Examination (M-ACE) was subsequently developed, which

appears to be more sensitive and specific than its widely used precursor, MMSE (Hsieh *et al.*, 2015; Folstein *et al.*, 2001).

The Edinburgh Cognitive and Behavioural ALS Screen (ECAS) was recently developed (Abrahams *et al.*, 2014) and adapted for the Greek (Kourtesis *et al.*, 2019), Italian (Poletti *et al.*, 2016), German (Lule *et al.*, 2015), Chinese (Ye *et al.*, 2016) and Spanish (Mora *et al.*, 2018) populations. ECAS is also a brief assessment similar to ACE-III, but it was designed for patients with various motor impairments and was found to be sensitive in amyotrophic lateral sclerosis (ALS), Parkinson's disease and progressive supranuclear palsy (Foley *et al.*, 2018; Niven *et al.*, 2015; Strong *et al.*, 2017). ECAS has been specifically designed to detect the type of cognitive and behavioural impairment in ALS of an executive nature similar to that found in FTD. ECAS comprises an ALS-specific component (executive function and social cognition, verbal fluency and language) and a carer's interview to detect the behavioural and psychotic changes typical in FTD. This focus on executive functions distinguishes ECAS from ACE-III. However, the ECAS was also designed to assess the functions typically affected in other diseases common in older adults, such as AD, and therefore it includes an ALS Non-Specific segment (memory and visuo-spatial function) (Foley et al., 2018; Niven et al., 2015; Strong et al., 2017). We have previously demonstrated that the ALS Non-Specific score is highly sensitive and specific in identifying the cognitive changes typical of AD and helps differentiate AD from ALS (Kourtesis et al., 2019).

In this study, aimed to adapt the ACE-III and M-ACE and to examine their convergent validity against their predecessors, ACE-R and MMSE, in a Greek population. Moreover, our primary aim was to compare these screening tools (ACE-III, M-ACE, ACE-R, MMSE and ECAS) in detecting AD in a Greek population.

Methods
Participants

All the participants and their carers signed an informed consent form in compliance with the revised Declaration of Helsinki (1987). This study was approved by the Psychology Research Ethics Committee of the University of Edinburgh, as well as the Aeginition Hospital Ethics Committee. All the participants were native Greek speakers and free from the following: (1) psychiatric disorders; (2) psychoactive drugs, antidepressants and anticonvulsants, (3) other neurological conditions affecting cognition; (4) learning disabilities; (5) alcoholism and drug abuse and (6) uncontrolled systemic diseases.

Patients with AD

The attendants of the Maroussi Alzheimer Clinic of the Athens Association of Alzheimer Disease and Related Disorders, Athens, Greece, were employed for this study. A total of 40 patients with AD participated; a sub-sample has previously been described by Kourtesis *et al.* (2019). Recruitment was conducted in accordance with the general inclusion criteria and the following criteria specific to AD: (1) a diagnosis of AD according to the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) (McKhann *et al.*, 1984) and (2) the absence of mixed concomitant dementia processes (e.g. AD and vascular dementia). In addition, a neuropsychologist or psychiatrist interviewed the patients and administered the Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983) (cut-off ≥ 8) to exclude patients with major depression or anxiety symptoms that may compromise their performance. The duration of the disease was calculated in years, from the onset of the first symptoms to the testing date.

Healthy Subjects

In this study, 38 controls were recruited and matched in age, sex and education to the patient group. They belonged to one of the following categories: (1) members of Athens Association of Alzheimer Disease and Related Disorders, Athens, Greece; (2) relatives of patients with AD or (3) volunteers who responded to the calls of the above association. For recruitment, we implemented the aforementioned general inclusion criteria.

Procedures

Translation-Adaptation

Adaptation of ACE-III and embedded M-ACE required minor adjustments as the administration (e.g. instructions) of the majority of the tasks was similar to that of ACE-R. The most significant adjustment, in terms of translation, was in the section of language in the task of proverb repetition, in which the pronunciation of proverbs is required. In terms of pronunciation, the first item should be a low-difficulty proverb (i.e. 'All that glitters is not gold'), and the second item should be a medium-to-hard-difficulty proverb (i.e. 'A stitch in time saves nine'). The proverbs of the Greek version were culturally adjusted, and the counterparts of this difficulty measure were 'All that glitters is not gold' and 'Better donkey-tying than donkey-seeking'. The original version of ACE-III (in which the M-ACE is embedded) was adapted to the Greek language using the back-translation method. The

original English version was translated into Greek by a native Greek speaker fluent in English, and then it was translated back into English by a native speaker of both Greek and English who was also blind to the original version of ACE-III. The procedure of translation/back translation was successful in only two iterations as there were mainly minor amendments compared to ACE-R. Finally, the adaptation of ECAS in Greek was described by Kourtesis *et al.* (2019).

Administration of the Tests and Inter-Rater Reliability

Administration of the tests was randomised to control for a possible practice effect (Benedict & Zgaljardic, 1998). Inter-rater reliability was calculated between the scores for ACE-III and M-ACE provided by the assessors and the independent reviewer. The four assessors and the independent reviewer were trained equally in the administration and scoring of ACE-III and M-ACE based on relevant guidelines. The independent reviewer was blinded to the identity of the examiner as well as the examinee.

Statistical Analyses

Bayesian statistics were preferred over null hypothesis significance testing (NHST). The Bayesian factor (BF_{10}) has been found to be more parsimonious than the p-value in evaluating evidence against H0 (Cox & Donnelly, 2011; Held & Ott, 2018; Wagenmakers et al., 2018a; Wagenmakers et al., 2018b). Importantly, the difference between BF_{10} and p-values is even greater (in favour of BF_{10}) in small sample sizes, which is pertinent to the present study (Held & Ott, 2018; Wagenmakers et al., 2018a; Wagenmakers et al., 2018b). A larger BF_{10} postulates more evidence in support of H1 (Cox & Donnelly, 2011; Held & Ott, 2018; Marsman & Wagenmakers, 2017; Wagenmakers et al., 2018a; Wagenmakers et al., 2018b). In this study, a threshold of $BF_{10} \geq 10$ was set for statistical inference, which postulates strong evidence in favour of H1 (Marsman & Wagenmakers, 2017; Wagenmakers et al., 2018a; Wagenmakers et al., 2018b), and corresponds to a p-value of < 0.01 (e.g. $BF_{10} = 10$) or to a p-value of <0.001 (e.g. BF10 > 11) (Cox & Donnelly, 2011; Held & Ott, 2018). However, we report both BF_{10} and p-values in this study. Finally, BF_{10} allows evidence in either direction (i.e. towards H1 and H0), and its measurement of evidence is insensitive to the stopping rule, which substantially mitigates the multiple comparisons problem and generates reliable and more generalisable results (Dienes, 2016; Marsman & Wagenmakers, 2017; Wagenmakers et al., 2018b).

The inter-rater reliability between the assessors who administered the screening procedures and the independent interviewer was appraised using the Intra-Class Correlation Coefficient (ICC), which displays outcomes from 'no match' = 0 to 'seamless match' = 1 (Weir, 2005). The internal consistency of the Greek ACE-III and M-ACE was determined by calculating Cronbach's alpha coefficient. A Cronbach's alpha coefficient of 0.70 or greater is considered substantial (Nunnaly, 1994). Demographic and cognitive data were analysed and compared. Shapiro–Wilk's test revealed non-significant results (i.e. normal distribution) for every variable. Between-group comparisons were made using Bayesian independent samples t-tests. The convergent validity of the screening tools was examined in the whole sample ($N = 78$, i.e. HC = 38 + AD = 40). The convergent validity and associations between the screening tools were probed and quantified using Bayesian Pearson's correlation analysis to ensure that our results are more reliable and generalisable. Receiver operating characteristic (ROC) curve analyses and area under the curve (AUC) were implemented to appraise the psychometric properties of the screening methods. All statistical analyses were performed using SPSS Statistics v.24.0 (scale, ROC and AUC analyses) (Release 2016; IBM Corp., Armonk, NY, USA) and JASP v.0.8.1.2 (Bayesian Pearson's correlation analyses, Bayesian independent samples t-tests) (JASP Team, 2017). Finally, a post hoc analysis (i.e. the achieved statistical power) of the Bayesian Pearson's correlations (i.e. the convergent validity of the screening methods) was performed using G*Power (Faul $et\ al.$, 2007; Faul $et\ al.$, 2009).

Results

Inter-Rater Reliability & Internal Consistency

The inter-rater reliability demonstrated an almost seamless agreement between the assessors, indicating substantial suitability for clinical measures (Weir, 2005). An ICC value of 0.92 was found for ACE-III and M-ACE (Weir, 2005). The scale analyses demonstrated excellent internal consistency of ACE-III and M-ACE with Cronbach's alpha = 0.79 (Nunnaly, 1994). We also inspected the internal consistency of ACE-III by replacing the repetition task of the culturally adjusted proverbs with the repetition task of phrases in ACE-R. The internal consistency of ACE-III with the repetition task of phrases (ACE-R) dropped to 0.77, indicating that the new repetition task of culturally adjusted proverbs contributed to the improvement of the internal consistency of ACE-III.

Convergent Validity

The Bayesian Pearson's correlation analyses robustly supported the convergent validity of ACE-III and M-ACE, as well as ECAS and its sub-scores, by indicating a large effect size (i.e. Pearson's r varied from 0.845 to 0.976), highly significant p-values (i.e. $p < 0.001$), highly extreme evidence of the Bayesian factor analysis (e.g. $BF_{10} = 1.299e+33$) and an almost perfect statistical power (i.e. $\approx100\%$). The statistics for the Bayesian Pearson's correlations are displayed in Table 1. ACE-III displayed a robust correlation with ACE-R. Equally, M-ACE substantially correlated with MMSE. Moreover, ECAS and its sub-scores significantly correlated with ACE-R.

Table 1 – Convergent validity: Bayesian Pearson's correlations

Correlational Pairs	Pearson's r	p-value	Statistical Power	BF_{10}
ACE-R & ACE-III	0.976 ***	p< .001	$\approx100\%$	6.009e+253
MMSE & M-ACE	0.863 ***	p< .001	$\approx100\%$	1.299e+33
ACE-R & ECAS Total Score	0.924 ***	p< .001	$\approx100\%$	2.172e+69
ACE-R & ECAS ALS-Specific	0.911 ***	p< .001	$\approx100\%$	2.589e+57
ACE-R & ECAS ALS Non-Specific	0.845 ***	p< .001	$\approx100\%$	1.286e+28

*BF= Bayes Factor; * $BF_{10} > 10$ ** $BF_{10} > 30$, *** $BF_{10} > 100$; For post hoc statistical power α < .001*

Sensitivity and Specificity in the Detection of AD

ROC and AUC analyses were executed to explore the psychometric properties of the screens in detecting AD. Figure 1 presents the ROC curves of each screen and sub-score. All the tests confirmed an adequately high level of sensitivity and specificity. Additionally, the analysis computed the sensitivity and specificity respective to different cut-offs, and the optimum cut-off to determine abnormality is shown (see Table 2). The ACE-III, ECAS, ACE-R, M-ACE, and ECAS-ALS Non-Specific covered the greatest AUC.

Figure 1- ROC curves: differentiation between patients with AD and controls
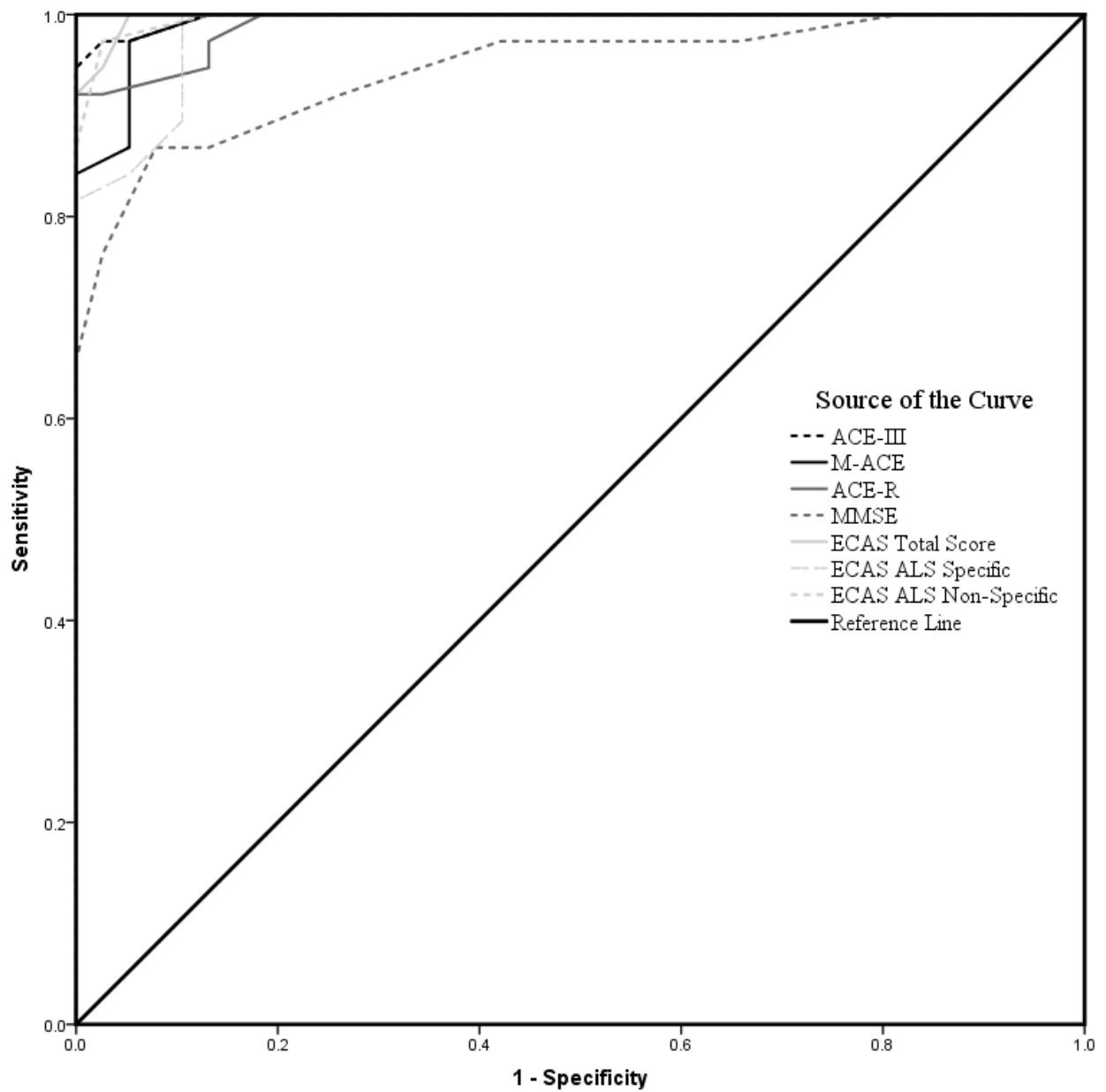
Table 2– Sensitivity and specificity in the detection of AD

Screen	AUC	PPV	NPV	PLR	NLR	Cut-off	Sensitivity	Specificity
ACE-III	99.7%	97.4%	97.4%	37.5	37.5	83.00	94.7%	100%
						84.00	**97.4%**	**97.4%**
M-ACE	99.1%	94.9%	97.3%	18.4	36.4	**23.00**	**97.4%**	**94.7%**
ACE-R	99.0%	100%	92.7%	92.1	12.5	82.00	89.5%	100%
						84.00	**92.1%**	**100%**
MMSE	94.5%	91.7%	87.5%	11.0	7.0	22.00	76.3%	97.4%
						24.00	**86.8%**	**92.1%**
ECAS	99.8%	97.3%	94.9%	36.4	18.4	93.00	92.1%	100%
						94.00	**94.7%**	**97.4%**
ECAS ALS Specific	98.4%	89.7%	91.9%	08.8	11.3	68.00	84.2%	94.7%
						71.00	**92.1%**	**89.5%**
ECAS ALS Non-Specific	99.7%	97.4%	97.4%	37.5	37.5	23.00	86.8%	100%
						24.00	**97.4%**	**97.4%**

The current cut-offs (2 SDs from the mean) are displayed first; The cut-ffs with highest sensitivity or specificity are presented. The proposed cut-offs (based on optimal sensitivity and specificity values) are showed in bold. Where the current and proposed cut-offs are the same only one value is given;
AUC = Area Under Curve; PPV = Positive Predictive Value; NPV = Negative Predictive Value;
PLR = Positive Likelihood Ratio; NLR = Negative Likelihood Ratio

Cognitive Performance and Behavioural Changes in AD

In the whole sample, there were no associations between cognitive performance and age/education. In group comparisons, patients with AD performed significantly worse than healthy controls in every test (see Table 3). In the AD sample, the HADS scores (i.e. depression and anxiety) did not correlate with cognitive performance. In addition, we examined the correlation between the screening methods and the duration of disease in the sample of patients with AD. Robust correlations with the duration of disease were detected solely with the total score of ECAS (BF_{10} = 14.22), whereas with the rest of the screening methods and sub-scores the correlations were non-significant (see Table 4). 14 patients (35%) had a disease duration of less than three years and 26 patients (65%) had a disease duration of three to six years, indicating that the sample of patients were in the early and early-middle stages of AD. Furthermore, the carers of 16 out of 40 patients with AD (40%) reported behavioural changes in the ECAS Behavioural Interview. The most prominent behavioural changes were apathy and loss of sympathy with some describing disinhibition (see Figure 2), whereas none of the carers reported a behavioural change pertaining to the rest of the ECAS behavioural items (i.e. compulsion, hyperorality and psychosis).

Table 3 – Comparison between controls and patients with AD

|	Controls – Mean (SD)	AD – Mean (SD)	p-value	BF_{10}
N = 78	38	40		
Sex	20M / 18F	19M / 21F	p = .646	(Chi^2 test)
Age	72.55 (6.32)	74.74 (6.05)	p = .128	0.103
Education	12.26 (3.20)	11.61 (3.25)	p = .377	0.531
ACE-III (Max Score = 100)	92.16 (4.08)	61.18 (16.86)	p < .001	***2.447e +14
M-ACE (Max Score = 30)	27.05 (2.16)	15.16 (6.06)	p < .001	***1.178e +15
ACE-R (Max Score = 100)	92.03 (3.82)	62.29 (17.03)	p < .001	***3.067e +13
MMSE (Max Score = 30)	27.53 (2.05)	19.37 (5.35)	p < .001	***2.475e +10
ECAS Total Score (Max Score = 100)	109.61 (8.30)	68.82 (18.08)	p < .001	***1.615e +17
ECAS-ALS Specific (Max Score = 100)	80.37 (6.26)	53.45 (14.03)	p < .001	***1.051e +14
ECAS-ALS Non-Specific (Max Score = 36)	29.24 (2.74)	15.37 (6.16)	p < .001	***1.953e +17

*SD = Standard Deviation; BF= Bayes Factor; * BF_{10} > 10, ** BF_{10} > 30, *** BF_{10} > 100*

Table 4 – Bayesian Pearson's correlations with the duration of disease

Correlational Pairs	Pearson's r	p-value	BF_{10}
Disease Duration & ACE-III	-0.307	p>.05	2.306
Disease Duration & M-ACE	-0.215	p = 0.91	0.839
Disease Duration & ACE-R	-0.338	p< .05	3.464
Disease Duration & MMSE	-0.108	p = 0.25	0.360
Disease Duration & ECAS Total Score	-0.424 *	p< .01	14.222
Disease Duration & ECAS ALS-Specific	-0.392	p< .01	7.995
Disease Duration & ECAS ALS Non-Specific	-0.348	p< .05	4.010

*BF= Bayes Factor; * $BF_{10} > 10$, ** $BF_{10} > 30$, *** $BF_{10} > 100$*

Figure 2- ECAS Behavioural Interview: behavioural changes in AD
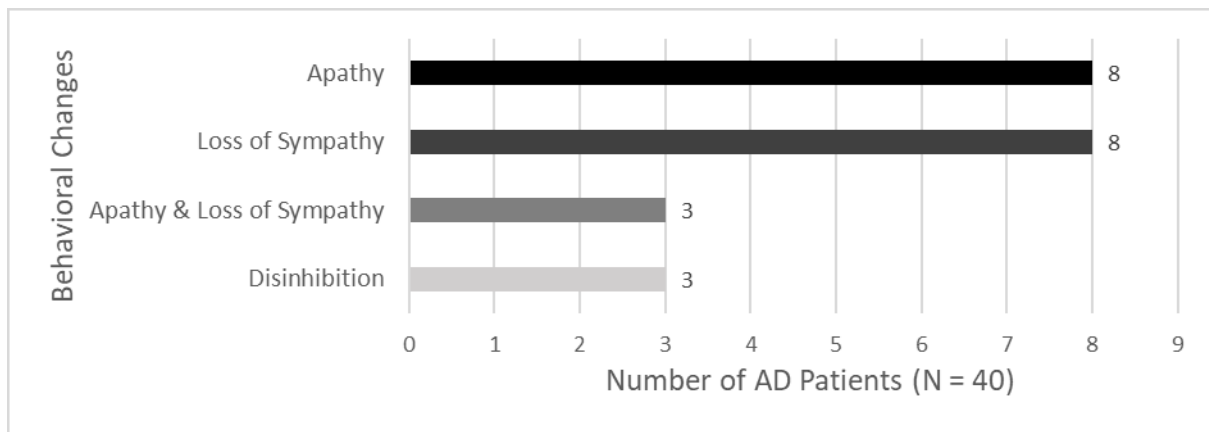
Ceiling Effects in the Screens

The ECAS scores appear to be substantially more resilient to ceiling effects compared to ACE-III, ACE-R (Figure 3) and, M-ACE and MMSE (Figure 4). Ceiling effects in the last two short screens were pronounced (see Figures 3 and 4). Lastly, only four patients with AD (out of 40, i.e. 10%) failed to collect two points (i.e. maximum points) in the phrase repetition task of ACE-R, whereas 10 patients with AD (i.e. 25%) failed to collect the maximum points in the proverb repetition task of ACE-III.

Figure 3- Distribution of healthy controls' performance in the fourth quartile of the possible scores
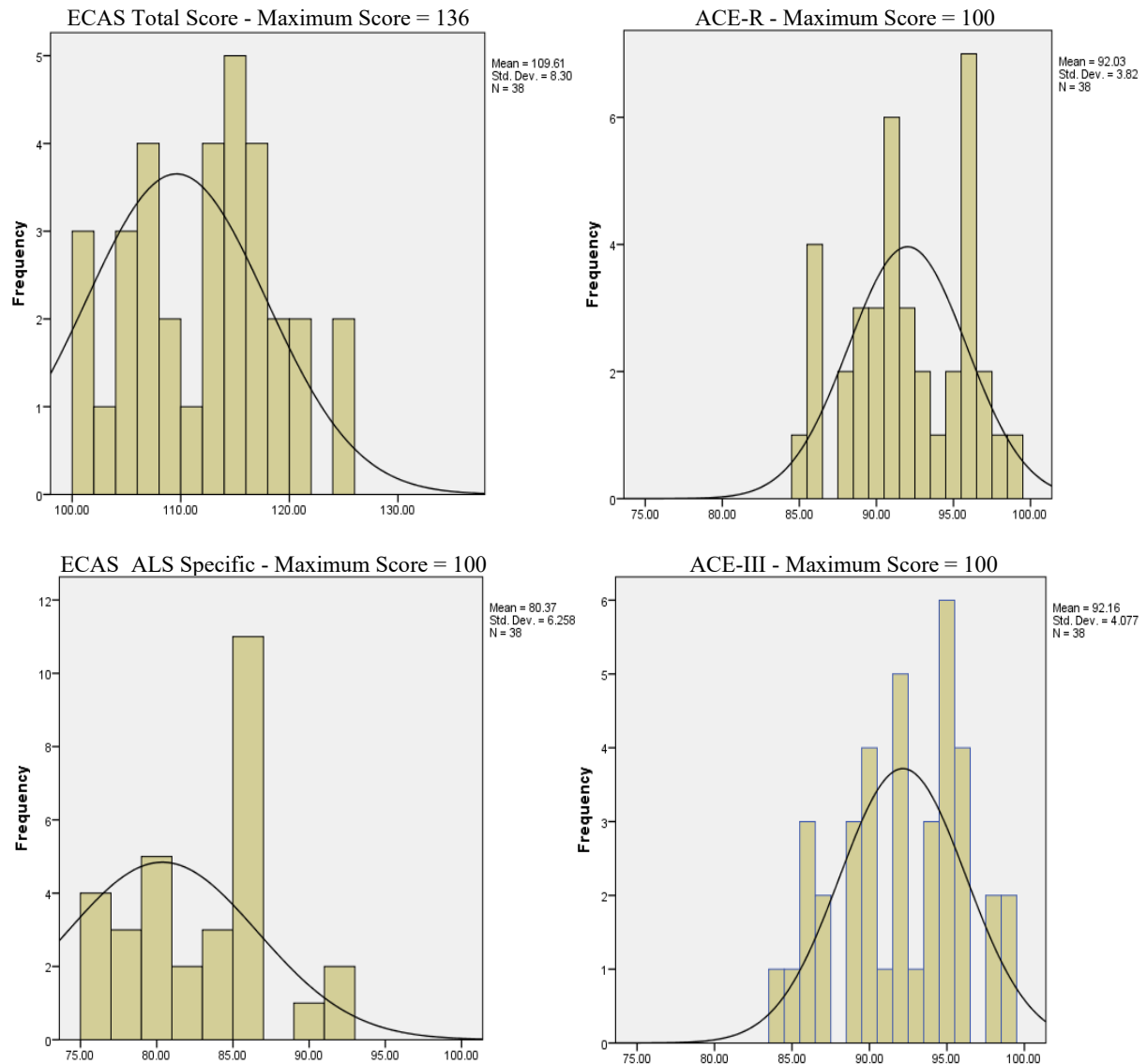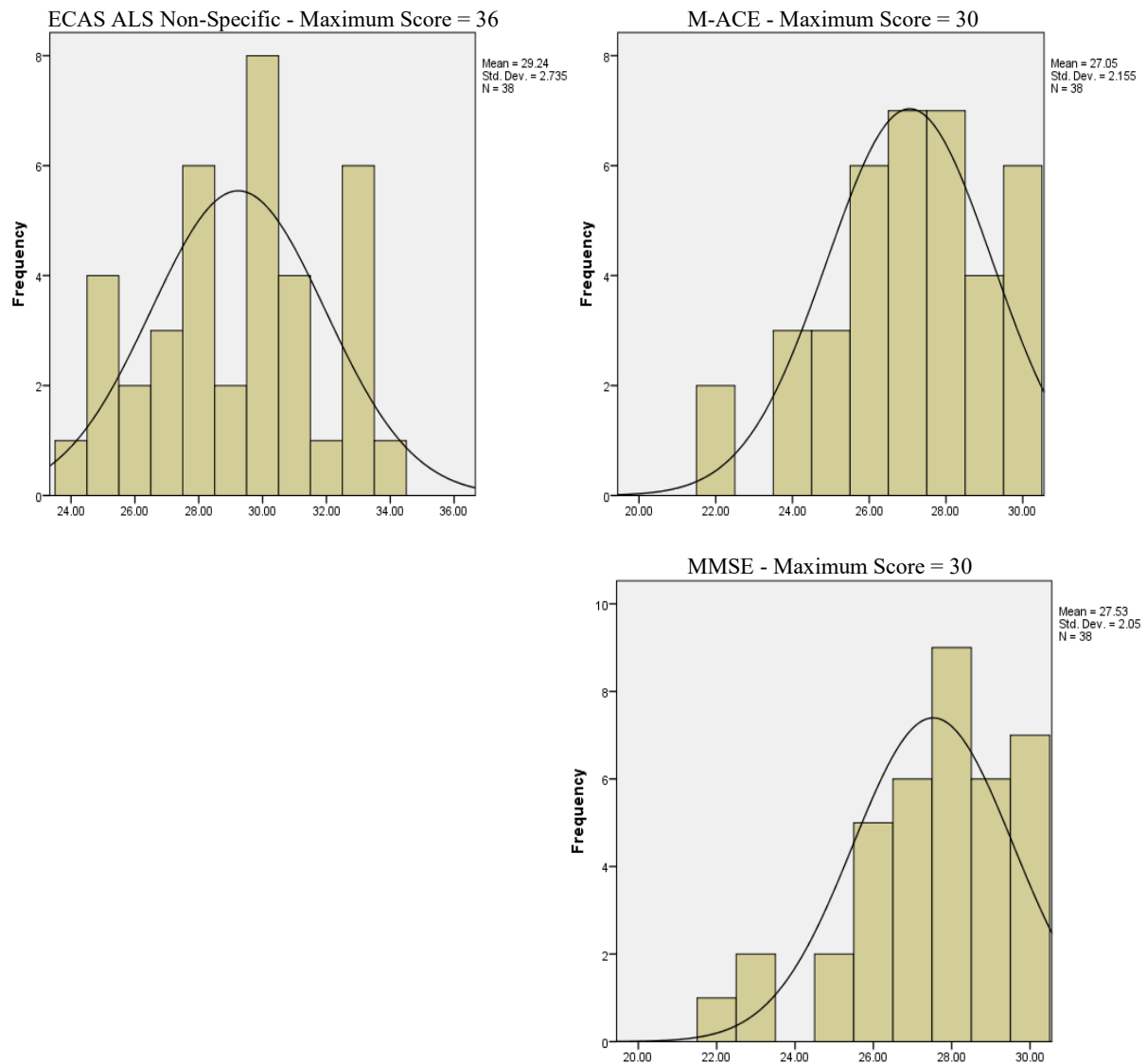

ECAS Total Score - Maximum Score = 136

Mean = 109.61
Std. Dev. = 8.30
N = 38

ACE-R - Maximum Score = 100

Mean = 92.03
Std. Dev. = 3.82
N = 38

ECAS ALS Specific - Maximum Score = 100

Mean = 80.37
Std. Dev. = 6.258
N = 38

ACE-III - Maximum Score = 100

Mean = 92.16
Std. Dev. = 4.077
N = 38

Figure 4- Distribution of healthy controls' performance in the third tier of the possible scores

Discussion

The present study successfully produced the Greek versions of ACE-III and M-ACE. The tests showed robust convergent validity against the already adapted and validated Greek versions of ACE-R and MMSE as evidenced by the large effect size, the high significance of the correlations, the strong evidence of the Bayesian factor analysis, and the almost perfect statistical power. Furthermore, the screening methods exhibited substantial internal consistency, which allows for implementation in clinical and research settings (Nunnaly, 1994). The tests also showed almost excellent inter-rater reliability, permitting extensive utilisation by various clinical practitioners (Weir, 2005). Therefore, the Greek ACE-III and M-ACE can be considered as suitable tools for clinical and research purposes.

ACE-III elicited 94.7% sensitivity and 100% specificity at a cut-off of 83 (2 SDs), as well as 97.4% sensitivity and 97.4% specificity at a cut-off of 84, in the detection of dementia within a sample pool of patients with AD who were predominantly in their first to fourth years after diagnosis. The sensitivity of ACE-III (94.7% and 97.4%) was superior to that of ACE-R (89.5% and 92.1%), demonstrating that ACE-III should be the tool of choice against ACE-R.

A comparison of M-ACE to MMSE revealed superior psychometrics in the former with 97.4% sensitivity and 94.7% specificity at a cut-off of 23 (MMSE, 86.8% sensitivity and 92.1% specificity at a cut-off of 24). The higher sensitivity and comparable specificity to MMSE are aligned with the validation study of M-ACE (Hsieh *et al.*, 2015). Accordingly, M-ACE surfaces as the most appropriate brief screening tool for detecting AD. M-ACE may be considered in hectic clinical environments, in which brief screening procedures are preferred.

Furthermore, the ECAS-ALS Non-Specific score was equally able to detect AD compared to the ACE-III with 97.4% sensitivity and specificity at a cut-off of 24. In addition, the ECAS-ALS Non-Specific score was substantially more specific than M-ACE but was equally sensitive. However, the sensitivity of the total score of ECAS was slightly below that of ACE-III and M-ACE although specificity was comparable. Of note the ECAS Total score was the only score that correlated with the disease duration, indicating that it may be more sensitive to cognitive decline than the rest of the screens, although this has yet to be demonstrated. Lastly, the ECAS-ALS Specific score appears to be less sensitive and specific compared to the above screening methods, although it displayed good psychometric properties in the identification of AD.

Utility of the Screens

The Greek version of ACE-III contains a repetition task of culturally adjusted proverbs, which replaced the repetition task of phrases in ACE-R. These items appeared to contribute to the internal consistency of ACE-III, and were less prone to ceiling effects in the AD group compared to the equivalent task of ACE-R. However, both tests suffered from ceiling effects in contrast to the ECAS. These were most pronounced in the shorter screening tools (M-ACE and MMSE, although, the former was marginally less prone than the latter, which is in line with the findings of Hsieh et al. (2015). However, the ECAS-ALS Non-Specific score and in particular the ECAS Total score and ECAS-ALS Specific did not suffer from ceiling effects. These findings are in line with the findings of a previous study in which ECAS was found to

be substantially less dependent on IQ and produced significantly fewer ceiling effects compared to ACE-III, which may be an advantage for use with clinical groups (De Icaza Valenzuela et al., 2018).

Furthermore, ACE-III does not include a behavioural assessment, which is a shortcoming (Hsieh et al., 2013). In contrast, the ECAS has a Behavioural Interview, which may add to the cognitive profile of the patient and predict caregiver's burden. In the current study, 40% of the carers of patients with AD reported behavioural changes. The most prominent were apathy and loss of sympathy. However, the ECAS Behavioural does not assess comprehensively apathy. Apathy is considered a multi-dimensional construct incorporating emotional, executive and initiation dimensions (Caga et al., 2018; Marin, 1991; Radakovic & Abrahams, 2018). Recently, the Dimensional Apathy Scale (DAS) was developed to assess these constructs (Radakovic & Abrahams, 2014). Notably, the DAS was implemented in AD, where a heterogenous profile emerged, enabling classification into three distinct groups (Radakovic et al., 2017). Hence, the DAS may be used in conjunction with ACE-III or ECAS in order to further identify and differentiate the types of apathy, which may be of clinical relevance.

Moreover, ACE-III and M-ACE are not adjusted to motor impairments, whereas ECAS is adjusted to upper motor and speech impairments. In a previous study, the ECAS-ALS Non-Specific score displayed very good psychometric properties in differentiating patients with AD from non-demented patients with ALS, whereas ACE-III and M-ACE were not successful (Kourtesis *et al.*, 2019). Therefore, the ECAS might be considered as an appropriate tool in patients with motor dysfunction, which are common in many neuro-degenerative diseases, and could be ideally included in future clinical trials.

Limitations & Future Studies

This study contains certain caveats that should be noted. One of the limitations of this study is the small sample size albeit the facilitation of robust statistical analyses with high statistical power. A larger and more diverse sample would allow more solid and conclusive observations. In future studies, the acquisition of normative data should be of a size that permits the computation of distinct cut-off scores that are analogous to the educational level.

Only patients with AD were recruited in this study. It would be of relevance to investigate the capacity of the tests to differentiate between patients with FTD and AD and probe FTD phenotypes. In addition, future studies should consider adapting a scale such as the DAS in

Greek, which may assist with research and/or clinical endeavours. The extensive and profound study of cognitive and behavioural changes in patients with dementia can help ameliorate and adjust patient care and alleviate the caregivers' burden.

Declaration

The authors declare no conflicts of interest and that this study is their own work.

References

Abrahams, S., Newton, J., Niven, E., Foley, J., & Bak, T. H. (2014). Screening for cognition and behaviour changes in ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 15(1-2), 9-14.

Benedict, R. H., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology*, 20(3), 339-352.

Brugnolo, A., Nobili, F., Barbieri, M. P., Dessi, B., Ferro, A., Girtler, N., ... & Servetto, G. (2009). The factorial structure of the mini mental state examination (MMSE) in Alzheimer's disease. *Archives of Gerontology and Geriatrics*, 49(1), 180-185.

Caga, J., Hsieh, S., Highton-Williamson, E., Zoing, M. C., Ramsey, E., Devenney, E., ... & Kiernan, M. C. (2018). The burden of apathy for caregivers of patients with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 19(7-8), 599-605.

Cox DR, Donnelly CA. (2011). *Principles of Applied Statistics*. Cambridge, UK: Cambridge University Press.

De Icaza Valenzuela, M. M., Bak, T. H., Pal, S., & Abrahams, S. (2018). The Edinburgh Cognitive and Behavioral ALS screen: relationship to age, education, IQ and the Addenbrooke's Cognitive Examination-III. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 19(7-8), 585-590.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78-89.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160.

Fields, J. A., Ferman, T. J., Boeve, B. F., & Smith, G. E. (2011). Neuropsychological assessment of patients with dementing illness. *Nature Reviews Neurology*, 7(12), 677.

Foley, J. A., Niven, E. H., Paget, A., Bhatia, K. P., Farmer, S. F., Jarman, P. R., ... & Abrahams, S. (2018). Sensitivity and Specificity of the ECAS in Parkinson's Disease and Progressive Supranuclear Palsy. *Parkinson's Disease*, 2018.

Folstein, M. F., Folstein, S. E., & Fanjiang, G. (2001). *Mini-mental state examination: clinical guide*. Lutz, FL: Psychological Assessment Resources.

Ganguli, M., Ratcliff, G., Huff, F. J., Belle, S., Kancel, M. J., Fischer, L., & Kuller, L. H. (1990). Serial sevens versus world backwards: a comparison of the two measures of attention from the MMSE. *Topics in Geriatrics*, 3(4), 203-207

Held, L., & Ott, M. (2018). On p-values and Bayes factors. *Annual Review of Statistics and Its Application*, 5, 393-419.

Hsieh, S., Hodges, J. R., Leyton, C. E., & Mioshi, E. (2012). Longitudinal changes in primary progressive aphasias: differences in cognitive and dementia staging measures. *Dementia and Geriatric Cognitive Disorders*, 34(2), 135-141

Hsieh, S., McGrory, S., Leslie, F., Dawson, K., Ahmed, S., Butler, C. R., ... & Hodges, J. R. (2015). The Mini-Addenbrooke's Cognitive Examination: a new assessment tool for dementia. *Dementia and Geriatric Cognitive Disorders*, 39(1-2), 1-11.

Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., & Hodges, J. R. (2013). Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 36(3-4), 242-250.

IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

JASP Team (2017). JASP (Version 0.8.1.2)[Computer software]

Kipps, C. M., Nestor, P. J., Dawson, C. E., Mitchell, J., & Hodges, J. R. (2008). Measuring progression in frontotemporal dementia: implications for therapeutic interventions. *Neurology*, 70(22), 2046-2052.

Konstantinopoulou, E., Kosmidis, M. H., Ioannidis, P., Kiosseoglou, G., Karacostas, D., & Taskos, N. (2011). Adaptation of Addenbrooke's Cognitive Examination-Revised for the Greek population. *European Journal of Neurology*, 18(3), 442-447.

Kourtesis P., Christidi F., Margioti E., Demenega C., Rentzos M., Evdokimidis I., Abrahams S. (2019). The Edinburgh Cognitive and Behavioural Amyotrophic Lateral Sclerosis Screen: Sensitivity in Differentiating between ALS and Alzheimer's Disease in a Greek Population. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 1-8

Lulé, D., Burkhardt, C., Abdulla, S., Böhm, S., Kollewe, K., Uttner, I., ... & Ludolph, A. C. (2015). The Edinburgh cognitive and behavioural amyotrophic lateral sclerosis screen: a cross-sectional comparison of established screening tools in a German-Swiss population. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 16(1-2), 16-23.

Marin, R. S. (1991). Apathy: a neuropsychiatric syndrome. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 3(3), 243-254.

Marsman, M., & Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, 14(5), 545-555.

Mathew, R., Bak, T. H., & Hodges, J. R. (2011). Screening for cognitive dysfunction in corticobasal syndrome: utility of Addenbrooke's cognitive examination. *Dementia and Geriatric Cognitive Disorders*, 31(4), 254-258.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7), 939-939.

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, 21(11), 1078-1085.

Mora, J. S., Salas, T., Fernández, M. C., Rodríguez-Castillo, V., Marín, S., Chaverri, D., & Rodríguez-Santos, F. (2018). Spanish adaptation of the edinburgh cognitive and behavioral amyotrophic lateral sclerosis screen (ECAS). *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 19(1-2), 74-79.

Niven, E., Newton, J., Foley, J., Colville, S., Swingler, R., Chandran, S., ... & Abrahams, S. (2015). Validation of the Edinburgh Cognitive and Behavioural Amyotrophic Lateral Sclerosis Screen (ECAS): a cognitive tool for motor disorders. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 16(3-4), 172-179.

Nunnally, J. C. (1994). Psychometric theory 3E. Tata McGraw-Hill Education.

Poletti, B., Solca, F., Carelli, L., Madotto, F., Lafronza, A., Faini, A., ... & Doretti, A. (2016). The validation of the Italian Edinburgh cognitive and behavioural ALS screen (ECAS). *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 17(7-8), 489-498.

Radakovic, R., & Abrahams, S. (2018). Multidimensional apathy: evidence from neurodegenerative disease. *Current Opinion in Behavioral Sciences*, 22, 42-49.

Radakovic, R., & Abrahams, S. (2014). Developing a new apathy measurement scale: Dimensional Apathy Scale. *Psychiatry research*, 219(3), 658-663.

Radakovic, R., Starr, J. M., & Abrahams, S. (2017). A novel assessment and profiling of multidimensional apathy in Alzheimer's disease. *Journal of Alzheimer's Disease*, 60(1), 57-67.

Raimondi, C., Gleichgerrcht, E., Richly, P., Torralva, T., Roca, M., Camino, J., & Manes, F. (2012). The Spanish version of the Addenbrooke's Cognitive Examination—Revised (ACE-R) in subcortical ischemic vascular dementia. *Journal of the Neurological Sciences*, 322(1-2), 228-231.

Strong, M. J., Abrahams, S., Goldstein, L. H., Woolley, S., Mclaughlin, P., Snowden, J., ... & Rosenfeld, J. (2017). Amyotrophic lateral sclerosis-frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 18(3-4), 153-174.

Valcour, V. G., Masaki, K. H., & Blanchette, P. L. (2002). The phrase:" no ifs, ands, or buts" and cognitive testing. Lessons from an Asian-American community. *Hawaii Medical Journal*, 61(4).

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... & Meerhoff, F. (2018a). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58-76.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... & Matzke, D. (2018b). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35-57.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research*, 19(1), 231-240.

Ye, S., Ji, Y., Li, C., He, J., Liu, X., & Fan, D. (2016). The Edinburgh cognitive and behavioural ALS screen in a Chinese amyotrophic lateral sclerosis population. *PloS one*, 11(5), e0155496.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67(6), 361-370.