Research Report

Reliability and Feasibility of the Memory Associative Test TMA-93

Emilio Franco-Macías[a,*], Silvia Rodrigo-Herrero[b], Andrea Luque-Tirado[a], Carlota Méndez-Barrio[b], Manuel Medina-Rodriguez[a], Eugenia Graciani-Cantisán[a], María Bernal Sánchez-Arjona[a] and Didier Maillet[c]

[a]*Unidad de Memoria, Servicio de Neurología, Hospital Universitario Virgen del Rocío, Seville, Spain*
[b]*Unidad de Memoria. Servicio de Neurología, Hospital Universitario Juan Ramón Jiménez, Huelva, Spain*
[c]*Service de Neurologie, Hôpital Saint-Louis (AP-HP), Paris, France*

Abstract.

Background: Memory tests focused on binding may be more sensitive to diagnose Alzheimer's disease (AD) at an early phase. TMA-93 examines relational binding by images.

Objective: Evaluate the reliability (internal consistency and inter-rater and test-retest reliability) and feasibility of the TMA-93 in a clinic setting with low-educated individuals and limited face-to-face time per patient.

Methods: The study was undertaken in a neurology outpatient clinic of a hospital in Southern Spain. The internal consistency of the TMA-93 was estimated in 35 patients with amnestic mild cognitive impairment (aMCI) and 40 healthy controls (HCs). The inter-rater reliability (by two raters) and feasibility (by recording the percentage of participants who completed the test, and by timing the administration time) were evaluated in HCs ($n = 16$), aMCI patients ($n = 18$), and mild dementia patients ($n = 15$). The test-retest reliability for the TMA-93 total score was studied in 51 HCs tested by the same examiner 2–4 months apart. The internal consistency was estimated by Cronbach's alpha. The inter-rater and test-retest reliability was quantified by the intraclass correlation coefficient (ICC). The administration time was compared by diagnosis.

Results: The internal consistency was "optimal" (Cronbach's alpha = 0.936). The test-retest reliability was "good" [ICC = 0.802 (CI 95% = 0.653–0.887)]. The inter-rater reliability was "optimal" [ICC = 0.999, (CI 95% = 0.999–1)]. All participants completed the test. The administration time ranged from less than 3 min in HCs to 6 min in aMCI patients, and 7 min in mild dementia patients.

Conclusion: Good feasibility and reliability support using the TMA-93 for examining visual relational binding, particularly in the context of low-educational attainment and limited time per patient.

Keywords: Binding, feasibility, inter-rater reliability, internal consistency, mild cognitive impairment, test-retest reliability, TMA-93

INTRODUCTION

In cognition, binding is the function that supports the integration of multiple elements together [1–3]. Errors in conjunctive binding (the integration of features within an object) and also in relational binding or associative memory (the ability to remember novel

*Correspondence to: Emilio Franco-Macías, Unidad de Memoria, Servicio de Neurología, Hospital Universitario Virgen del Rocío, Avenida Manuel Siurot s/n, Seville 41013, Spain. Tel.: +34 609732041; Fax: +34955012593; E-mail: efranco17@gmail.com.

associations between words or pictures) have been reported in Alzheimer's disease (AD) at an early phase [4]. Conjunctive binding is supported by the entorhinal and perirhinal cortex and seems more sensitive than the relational one to early AD [5, 6]. The "Short-Term Memory Binding Test" is the most used tool for examining conjunctive binding. This test, one of the most promising neuropsychological tools, has been incorporated into trials to predict who among those with mild cognitive impairment will go on to develop AD [7]. There is also evidence that relational binding, that relies on the hippocampus, parahippocampal cortex, and default mode network regions (posterior cingulate cortex/precuneus/lateral parietal and medial frontal cortex) [5], declines in the prodromal stages of late-onset sporadic AD [8–10]. Even more, asymptomatic individuals with greater amyloid-β burden on amyloid imaging have shown abnormal scores on relational binding tests when the performance on other standardized episodic memory test is still preserved [11]. In neuropsychology, the relational binding ability can be examined by different tests. The "Wechsler Memory Scale" (WMS) assesses binding through learning and recall of paired associated words [12]. This WMS subtest discerns between easy (i.e., North/South) and complex associations (i.e., School/Cellar) [12]. The "Memory Binding Test" (MBT) examines associative memory through the recall of pairs of items that belong to the same semantic category (i.e., flea/ant = insects) but presented in two different lists of words [13]. The "Face Name Associative Memory Exam" is a cross-modal associative test based on a functional magnetic resonance imaging (fMRI) task that pairs pictures of unfamiliar faces with common first names [14].

Testing relational binding only by images rather than words could be more feasible for low educated individuals. The "Memory Associative Test of the district of Seine-Saint-Denis" (TMA-93) was recently developed in France for the early diagnosis of AD among low educated immigrants [15]. Briefly, during the encoding phase, the patient is shown ten pairs of drawings of common and easy to recognize objects from daily life that are semantically related (Fig. 1A). Only one of the two items is shown in the recall phase, and the patient is asked to recall the missing item (Fig. 1B) [15]. In the original paper, the test demonstrated high diagnostic accuracy for discriminating AD patients from healthy controls in a sample of immigrant residents from a district in Paris (France) [15]. In that study, the cutoff of 24 of 30 showed a sensitivity of 88% and a specificity
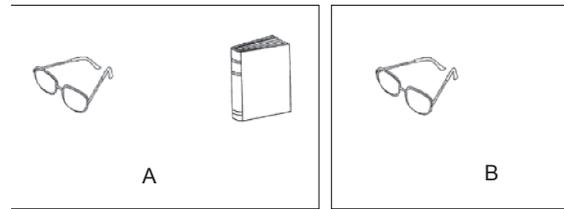

Fig. 1. Pairs of semantically-related drawings of the TMA-93. In the codification phase, the semantically-related drawings are presented in pairs (A). In the recall phase, the subject has to recall the missing object (B).

of 97% for distinguishing AD patients from healthy controls [15]. A posterior validation study in older educationally-diverse Spanish people demonstrated that the test is so sensitive as the picture version of the Free and Cued Selective Reminding Test (FCSRT) in discriminating between amnestic mild cognitive impairment (aMCI) patients and healthy controls (HCs) [16]. In that study, the receiver operating characteristic (ROC) curve analysis determined an optimal area under the ROC curve (AUC) of 0.97 (95% CI, 0.89–1.00, $p < 0.001$) to distinguish between aMCI patients and HCs [16].

On the other hand, the most used memory tests are based on learning and recalling two paragraphs or a list of words and often include a final step of facilitation with verbal cues or recognition among distractors that have to be administered 15–30 min later [17, 18]. These tests take too long time to be used in busy primary care and general neurology outpatient settings with limited face-to-face time per patient. On the contrary, the TMA-93 is a relatively short test that may be more suitable in that context.

These potential uses and advantages of the TMA-93 encourage the completion of the development of the test. There are no previous studies focused on the reliability or feasibility of the TMA-93. There is a need to validate the test-retest reliability of binding tasks to detect and monitor AD-related populations [4]. Tests providing such reliability will be appropriate for use in longitudinal research. On the other hand, feasibility has been considered a crucial prerequisite by a consensus document on neuropsychological assessment [19]. The aim here was to study the reliability (the internal consistency and the inter-rater and test-retest reliability) of the TMA-93 and its feasibility (by recording the percentage of participants who completed the test, and by timing the administration time).

METHODS

Study population

The studies were undertaken in a general neurology outpatient clinic at the University Hospital Virgen del Rocio, a tertiary referral academic center in Seville, in the Southern Spanish region of Andalusia. Many older people of this region had limited access to primary school and are not skilled in reading or writing. In the region, time availability for examining patients with memory complaints is limited: from 5 min in busy primary care to 20 min in a general neurology outpatient setting.

The internal consistency was studied in an extension of the phase I validation study for the TMA-93 [16]. Here, the sample was increased to 75 individuals (35 patients with aMCI and 40 HCs) to meet the required sample size for studying the internal consistency of a test composed of 10 items. Procedures for this cross-sectional study has been previously described [16] and included Phototest, a brief cognitive test developed in Spain with high diagnostic accuracy for diagnosing cognitive impairment and dementia [20], and the "Delayed Matching-to-Sample Task 48" (DMS-48), a visual recognition memory task on which the diagnosis of aMCI was based [21]. The diagnosis of aMCI had been made according to the National Institute on Aging and Alzheimer's Association (NIA-AA) recommendations [22] and operationally put into practice as follows: 1) memory complaint corroborated by a reliable informant; 2) objective memory impairment measured by a score equal to or below the 10‑percentile on set 2 of DMS-48 (this score being lower than that on set 1); and 3) no significant functional decline for activities of daily living [score up to 39 on "Interview for Deterioration in Daily Living Activities in Dementia" (IDDD) (23)]. We recruited HCs among the caregivers and relatives of patients attending the center. They met the following inclusion criteria: 1) absence of memory complaints; 2) absence of objective memory impairment (DMS-48 set 2 score equal to or above the 25-percentile); and 3) intact level of independence in activities of daily living (score between 33 and 36 on IDDD).

For the test-retest reliability, HCs were recruited among the participants in the Spanish normative study for the TMA-93 [24]. The inclusion criteria for this study were: 1) age equal to or above 50; 2) no cognitive complaints; 3) score equal or above 10-percentile according to normative data for the Phototest in Spain [25]; and 4) independent level of functioning. 51 randomly selected HCs were invited to repeat the TMA-93 conducted by the same examiner (SRH) between 2 and 4 months after the initial examination.

For studying the inter-rater reliability and feasibility, an ad-hoc sample composed of 16 HCs, 18 patients with aMCI, and 15 patients with mild dementia due to probable AD ($n = 15$) was recruited. Both groups of patients had been diagnosed according to NIA-AA recommendations [22, 26]. The diagnosis of mild dementia due to probable AD was based on core clinical criteria for AD [26]. Amnestic presentation and classification at stage 4 according to the Global Deterioration Scale were required [27]. All available information had been used for this diagnostic process including history, blood tests, brain imaging (head CT or brain MRI), and the following battery of neuropsychological tests: the Spanish version of the Informant Questionnaire AD8 [28], the Phototest [20], the picture version of the FCSRT [29], the Stroop Color and Word Test [30], the ADAS-Cog subtest of constructive praxis [31], the 12-item Boston Naming Test [32], the VOSP subtests of Dot Counting and Number Location [33], the IDDD [23], and the 15-item Geriatric Depression Scale [34]. The TMA-93 was conducted by two examiners (1 = EGC and 2 = SRH), who followed an alternating order for its administration and timing on the same subjects, and were blinded to both the subject's diagnosis and the score obtained by the other examiner.

Instrument: TMA-93

The TMA-93 was administered following the instructions given by its authors [15]. During the encoding phase, subjects were shown and asked to name 10 pairs of real-life semantically-related objects presented as drawings in cards (tree/bird, bed/bedside lamp, boat/fish, dog/sheep, foot/trousers, knife/apple, glasses/book, hand/watch, car/car keys, flower/sun). The examiner specifically asked the participants to memorize the pairs of drawings (Fig. 1A). Next, the first associative memory trial was administered: examinees were shown only one of each pair's drawings and asked to recall the missing one (Fig. 1B). After each subject's response (regardless accuracy) or a period of up to 5 s, we displayed the pair again. This protocol was repeated for the 9 remaining pairs.

The maximum score of 30 points was granted only when the participant produced 10 out of 10 correct

responses in this first trial, in which case, the second and the third trials were omitted. Otherwise, the participants were scored from 0 to 9 based on their number of correct answers in this first trial and were administered a second similar trial with the same 10 pairs of drawings. If a subject correctly recalled the 10 missing objects in this second trial, s/he was given 20 points: 10 points corresponding to the second trial, and 10 more corresponding to the third trial, which was cancelled. The score of each of the 10 items of the TMA-93 ranged from 0 to 3 and these scores were used for estimating the test's internal consistency.

Three types of incorrect answers were recorded: 1) error, when the subject recalls an object that belongs to a different pair; 2) intrusion, when the subject recalls an object that was never shown to him/her; and 3) perseveration, when the subject repeated the same error [15].

Ethics

The studies were approved by the ethics committee of the Hospital Virgen del Rocio (Seville, Spain) and conducted according to the World Medical Association Declaration of Helsinki. All participants accepted the study procedures by signing informed consent.

Statistical analyses

Descriptive results are shown as frequency (percent) for dichotomous and categorical variables, mean ($\pm$SD, range) for normally-distributed continuous variables, and median [interquartile range (IQR), range] for non-normally distributed continuous variables. Between-group comparisons of continuous variables were performed with Student's *t*-test or one-way ANOVA (or their non-parametric alternatives Mann-Whitney U test and Kruskal-Wallis ANOVA, respectively). Between-group comparisons of categorical variables were performed with the Chi square test.

Internal consistency was estimated by Cronbach's alpha. Values of Cronbach's alpha above 0.70 were considered acceptable, between 0.90 and 0.95 were considered "optimal", and above 0.95 were interpreted as indicative of "item redundancy" [35, 36]. In addition, "split-half reliability" was analyzed considering the first five pairs of drawings of the TMA-93 as a half and the last five ones as the other half and estimating the correlation between each other by the

Spearman-Brown coefficient. "Corrected item-total correlations" were calculated, and a value below 0.40 was considered indicative of item redundancy [35]. Item redundancy was also evaluated by "Cronbach's alpha if item deleted", considering an item as redundant if the Cronbach's alpha increased at deleting it [37].

Test-retest reliability for the TMA-93 total score was estimated by the intra-class correlation coefficient (ICC). In addition, we also created the variable "total score time 2 minus total score time 1" and analyzed its distribution.

Inter-rater reliability for the TMA-93 total score and number of errors, intrusions, and perseverations were estimated by the ICC.

According to the ICC, reliability was categorized as: optimal (ICC > 0.90), good (ICC 0.71–0.90), moderate (ICC 0.51–0.70), mediocre (ICC 0.31–0.50), or bad/null (ICC < 0.31) [21].

The feasibility was analyzed by recording the number of participants who completed the test, and comparing the administration time according to diagnosis, and educational attainment.

Statistical significance was set at a $p < 0.05$, and all estimates were obtained with a 95% confidence interval (CI 95%).

All statistical analyses were run in SPSS version 25 (IBM, USA).

RESULTS

Socio-demographics characteristics and neuropsychological background for the extension of the cross-sectional study focused on internal consistency are shown in Table 1. For the total sample ($n = 75$), 46 participants were females. Their average age was 74.6 (SD = 5.9, range = 51–84). Regarding educational attainment, 31 individuals (41.3%) had not completed primary studies (Table 1). There were no significant differences in age, gender, or educational attainment between aMCI and HCs groups (Table 1). aMCI patients scored significantly lower than HCs on Phototest, DMS48, and TMA-93 (Table 1). Internal consistency was "optimal" (Cronbach's alpha = 0.936). Split-half reliability was also high (Spearman-Brown coefficient = 0.911). Corrected item-total score correlations ranged from 0.661 for the pair "hand-watch" to 0.837 for the pair "flower-sun" (Table 2). There was no redundancy of any item as the Cronbach's alpha did not increase at deleting anyone (Table 2).

Table 1
Socio-demographics characteristics and neuropsychological background of the internal consistency study

	Total, $n = 75$	HCs, $n = 40$	aMCI, $n = 35$	p
Age	74.6 ± 5.9 (51–84)	74.7 ± 6.3 (51–83)	74.6 ± 5.4 (65–84)	0.706
Gender				
Female	46 (61.3%)	21 (52.5%)	25 (71.4%)	0.093
Male	29 (38.7%)	19 (47.5%)	10 (19.6%)	
Educational attainment				
<first grade	31 (41.3%)	12 (30%)	19 (54.3%)	0.052
First grade	19 (25.3%)	14 (35%)	5 (14.3%)	
>first grade	25 (33.3%)	14 (35%)	11 (31.4%)	
Phototest (total score)	31.8 ± 7.5 (13–52)	36.3 ± 5.7 (26–52)	27.1 ± 6.3 (13–41)	<0.001
DMS48				
Set 1 score	45, (41–47), (31–48)	47, (46–47), (41–48)	41, (35–44), (31–47)	<0.001
Set 2 score	43, (36–47), (26–48)	47, (45–48), (40–48)	36, (30–39), (26–45)	<0.001
TMA-93 (total score)	24, (14–29), (0–30)	29, (25–30), (14–30)	13, (6–20), (0–28)	<0.001

Results are shown as median, (interquartile range), and (range) for non-normal distributed variables and mean ± SD, and (range) for normal distributed variables.

Table 2
Corrected Item-Total Correlations and Cronbach's alpha if item deleted

	Item-Total Statistics			
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
tree/bird	18,6667	67,793	0,773	0,927
bed/bedside lamp	18,4933	69,199	0,773	0,928
boat/fish	18,5067	68,929	0,744	0,929
dog/sheep	18,4800	71,253	0,677	0,932
foot/trousers	18,7200	68,366	0,722	0,930
knife/apple	18,5867	68,921	0,770	0,928
glasses/book	18,9333	68,441	0,719	0,930
hand/watch	18,6267	70,940	0,661	0,933
car/car keys	18,6400	70,098	0,754	0,929
flower/sun	19,1867	65,262	0,837	0,924

Corrected Item-Total Correlation was never lower than 0.400. Cronbach's alpha was not above 0.936 at deleting any item. Both results demonstrated no redundancy of any item.

Socio-demographics characteristics and neuropsychological background for the test-retest reliability study are shown in Table 3. Their average age was 64.8 (SD = 8.9, range = 50–86). 13 subjects (25.5 %) had not completed primary studies, and 30 (58.8%) were females (Table 3). Test-retest reliability for the TMA-93 total score was "good" [ICC = 0.802 (CI 95% = 0.653–0.887)]. The "total score time 2 minus total score time 1" variable showed a non-normal, right asymmetric, and leptokurtic distribution (median = 0, IQR = 0–1, Range = –3–3). There were four atypical observations: two of them scored three points higher at the retest and the remaining two scored two and three points lower, respectively (Fig. 2). We analyzed the TMA-93 total score at time 2 by the TMA-93 total score at time 1: the variability was greater for scores below 28, and some practice effect could be detected in the range 27–29 (Fig. 3).

Table 4 shows the characteristics of the sample for the inter-rater reliability and feasibility study. Their

Table 3
Socio-demographic characteristics and neuropsychological background of the test-retest study

Age	64.8 ± 8.9 (50–86)
Gender	
Female,	$n = 30$ (58.8%)
Male,	$n = 21$ (41.2%)
Educational attainment	
<first grade,	$n = 13$ (25.5%)
First grade,	$n = 16$ (31.4%)
>first grade,	$n = 22$ (43.1%)
Phototest (total score)	37.8 ± 4.8 (27–47)
TMA-93 (total score)	29, (28–30), (23–30)

Results are shown as median, (interquartile range), and (range) for non-normal distributed variables and mean ± SD, and (range) for normal distributed variables.

average age was 68.7 (SD = 7.2, range = 55–81). 16 subjects (32.7%) had not completed primary studies, and 32 (65.3%) were females. There were statistically significant differences in the TMA-93 scores across the three diagnostic groups
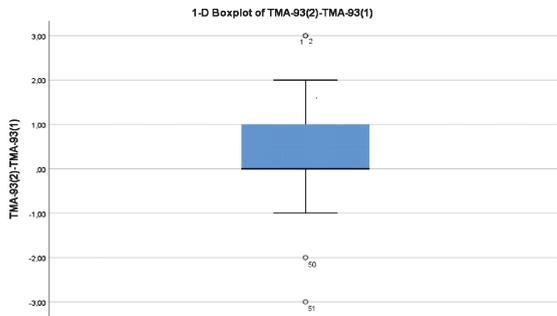
Fig. 2. Boxplot chart showing the distribution of the "total score time 2 minus total score time 1" variable. There are four outliers.
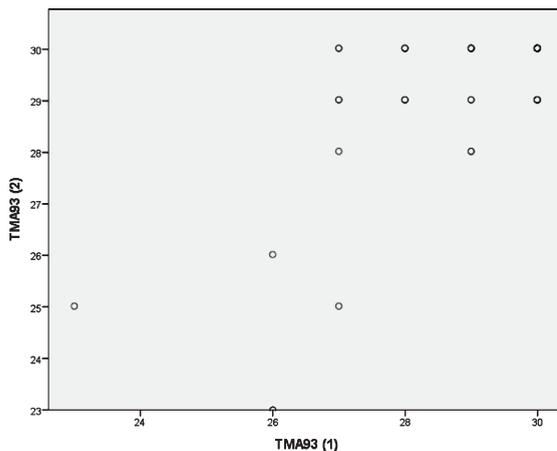

Fig. 3. Scatterplot with Time 1 performance on the x axis and Time 2 performance on the y axis. The variability in the measure was greater for scores below 28, and some practice effect could be detected in the range 27–29.

Fig. 4. Boxplot chart depicting the differences in "Administration Time" (in minutes) among the three diagnostic groups of the inter-rater reliability and feasibility study.

Fig. 5. Boxplot chart depicting no significant differences in "Administration Time" (in minutes) by educational attainment in the inter-rater reliability and feasibility study.

(Table 4). The inter-rater reliability was "optimal" for the TMA-93 total score [ICC = 0.999, (CI 95% = 0.999–1)], number of intrusions [ICC = 0.985 (CI 95% = 0.974–0.992)], and number of errors [ICC = 0.996 (CI 95% = 0.993–0.998)]. The inter-rater reliability for the number of perseverations was "good" [ICC = 0.853 (CI 95% = 0.738–0.918)].

All participants, including mild AD dementia patients, completed the test. There were statistically significant differences in the TMA-93 duration across the three diagnostic groups (Table 4). *Post-hoc* multiple comparison analyses revealed that the duration of the administration (in minutes) was significantly lower in healthy controls (median = 2.2, IQR = 2.0–4.0, range = 1.5–5.5) than in aMCI (median = 6.2, IQR = 4.7–7.8, range = 2.3–11.7, $p < 0.05$) and mild AD dementia patients (median = 7.5, IQR = 5.9–9.4, range = 5.0–17.2, $p < 0.001$). The
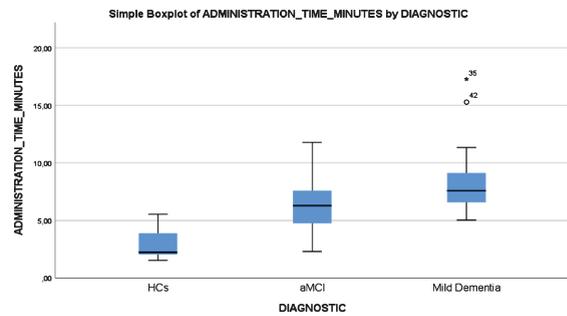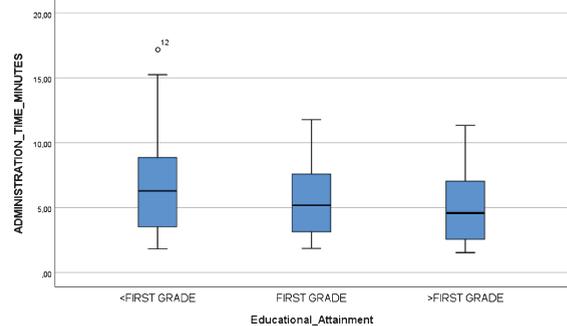
aMCI and mild AD dementia groups did not differ significantly ($p = 0.337$). There were two outliers from the mild dementia group, with an administration time longer than 15 min (Fig. 4). There were no statistically significant differences in the TMA-93 administration time by educational attainment in the inter-rater reliability study (<first grade: median = 6.28, IQR = 2.94–9.00, range = 1.82–17.25; first grade: median = 5.18, IQR = 2.63–7.58, range = 1.85–11.78; >first grade: median = 4.58, IQR = 2.43–7.10, range = 1.53–11.33; $p = 0.399$) (Fig. 5). To better analyze the educational attainment effect on the administration time, we went back to the test-retest study and evaluated differences in administration time by educational attainment among the HCs at test 1. Again, there were no significant differences (<first grade: median = 2.47, IQR = 1.77–3.40, range = 1.37–4.59; first grade: median = 2.23, IQR = 1.44–2.46, range = 1.38–3.16; >first grade: median = 2,26, IQR = 1.88–3.15, range = 1.46–4.11; $p = 0.352$) (Fig. 6).
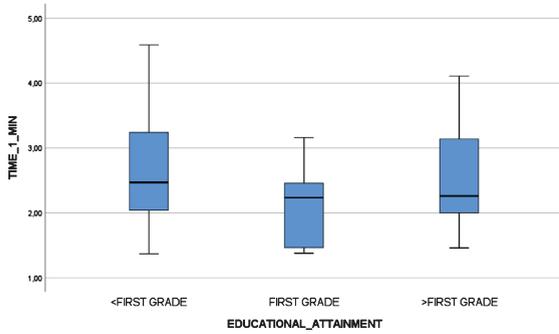
Fig. 6. Boxplot chart depicting no significant differences in "Administration Time" (in minutes) by educational attainment in the test-retest reliability study at time 1.

DISCUSSION

To our knowledge, this is the first study focused on the reliability and feasibility of the TMA-93, the French visual relational binding test [15]. The test has already demonstrated high diagnostic accuracy in validation studies [15, 16] and has normative studies from French and Spanish populations [24, 38].

Internal consistency among the 10 pairs of semantically-related drawings of real-life objects that compose the TMA-93 was "optimal" (Cronbach's alpha = 0.936). This result means the 10 items of the test are highly correlated each other and measure the construct of interest, visual relational binding, in a

similar way [39]. By comparison, an "acceptable" internal consistency has been reported for the FCSRT, a standard memory test (Cronbach's alpha = 0.810) [40].

"Corrected Item-Total Correlation" is the correlation of the item designed with the summated score for all other items. A rule-of-thumb states that this value should be at least 0.40 to rule out item redundancy [35]. Every item of the TMA-93 fulfilled the rule. In the same way, the Cronbach's alpha did not increase at deleting any of the ten pairs, so, again, redundancy of any item could not be demonstrated.

Split-half testing is another measure of internal consistency. This method measures the extent to which all parts of the test contribute equally to what is being measured. We found a strong correlation between the two virtual halves of the TMA-93, indicating that HCs and aMCI patients performed equally well (or as poorly) on both halves of the test.

The TMA-93 showed a "good" test-retest reliability [ICC = 0.802 (CI 95% = 0.653–0.887)]. By comparison, this reliability is similar to that reported for the "Mini-Mental State Examination" (MMSE) (0.80) [41] and suggests stability in performance over time. The test-retest reliability studies' design varies by the time considered between test 1 and test 2, and by the selection of participants (only HCs or mixed sample of HCs and patients). Here, we considered 2–4 months for administering the retest and only HCs. The time interval seems to be short enough to pre-

Table 4
Characteristics of the sample for the inter-rater reliability and feasibility study

	Healthy Controls	aMCI	Mild dementia due to AD	p
N	16	18	15	
Age	66,6 ± 6,4 (56–75)	69,8 ± 6,4 (58–80)	69,7 ± 8,8 (55–81)	0.38
Gender (F/M)	12 (75%)/4 (25%)	10 (55,5%)/8(45.5%)	10 (66,6%)/5 (33.3%)	0.48
Education				
<first grade	4/16 (25%)	7/18 (38.8%)	5/15 (33.3%)	0.93
First grade	7/16 (43.7%)	7/18 (38.8%)	6/15 (40%)	
>first grade	5/16 (31.2%)	4/18 (22.2%)	4/15 (22.2%)	
Duration of test (min)	2.2, (2.0–4.0), (1.5–5.5)	6.2, (4.7–7.8), (2.3–11.7)	7.5, (5.9–9.4), (5.0–17.2)	<0.001
TMA-93 (1)	30, (28–30), (24–30)	20, (6–27), (4–30)	6, (4–19), (0–24)	<0.001
TMA-93 (2)	30, (28–30), (24–30)	20, (6–27), (4–30)	6, (4–19), (0–24)	<0.001
Errors (1)	0, (0–0), (0–0)	0, (0–1), (0–12)	1, (0–2), (0–6)	<0.005
Errors (2)	0, (0–0), (0–0)	0, (0–1), (0–11)	0, (0–2), (0–7)	<0.01
Perseverations (1)	0, (0–0), (0–0)	0, (0–1), (0–2)	0, (0–1), (0–8)	0.074
Perseverations (2)	0, (0–0), (0–0)	0, (0–1), (0–2)	0, (0–1), (0–6)	0.075
Intrusions (1)	0, (0–0), (0–1)	2, (0–3), (0–18)	0, (0–3), (0–12)	<0.05
Intrusions (2)	0, (0–0), (0–1)	2, (0–3), (0–18)	0, (0–2), (0–13)	<0.05

aMCI, amnestic mild cognitive impairment; TMA-93 (1); TMA-93 total score by examiner 1; TMA-93 (2); TMA-93 total score by examiner 2; Errors (1) errors score by examiner 1; Errors (2), errors score by examiner 2; Perseverations (1), perseverations score by examiner 1; Perseverations (2), perseverations score by examiner 2; Intrusions (1), intrusions score by examiner 1; Intrusions (2), intrusions score by examiner 2. Age is expressed in mean ± SD and (range). Scores and duration of the TMA-93 are expressed in median, (interquartile range), and (range).

vent the effect of an eventual cognitive impairment on the sample, particularly from participants with lower scores, and long enough to prevent a practice effect. With a similar design, the MBT demonstrated ICC values ranged from 0.64 to 0.76 [42]. Analyzing the distribution of the "total score time 2 minus total score time 1" variable, there were four atypical observations that probably precluded this reliability could be upgraded to "optimal". Two of them scored 3 points more at the retest. On the opposite side, two outliers scored 2 and 3 points less, respectively. The former could be explained by practice effect and the latter by cognitive decline, but a more global explanation could be that binding is somewhat changeable and dynamic, making it difficult for a test to achieve an "optimal" test-retest reliability [43]. The variability in the measure were greater for scores below 28 at time 1. The test-retest reliability could be supported by scores above 28 at time 1 and, thus, overestimated due to ceiling effect. To clarify this issue, future test-retest reliability TMA-93 studies should recruit enough HCs scoring below 28 at time 1 and consider participants' AD biomarker status to understand eventual score changes over time.

Inter-rater reliability of the TMA-93 was "optimal" for the total score and the number of errors and intrusions, and "good" for the number of perseverations. We noted that the administration and scoring are relatively simple, but that classifying the incorrect responses in errors, intrusions or perseverations can lead to disagreements between examiners and requires some training. Individually, perseverations—scored as the number of times that an error (a response that corresponds to a different drawing pair) is repeated—were the main source of disagreement between examiners.

Regarding TMA-93 feasibility, all participants, including mild AD dementia patients, were able to complete the test. Participants' task-tolerability was good, including that of those who scored the minimum (4 out of 30) or whose administration time was the longest (17.2 min). There were significant differences in administration time by diagnosis: cognitively impaired patients spent more time on recalling the missing drawing, made more mistakes, and usually needed the maximum of three memory trials.

The average time required to complete the test was 2–3 min for HCs, 6 min for aMCI patients, and 7 min for mild AD dementia patients, so this test is relatively short despite being a specific memory test and not a brief cognitive screening test as MMSE or MoCA. By comparison, the time of passing the MMSE in cognitively impaired patients is, on average, 4 min 51 s [44]. Busy primary care and general neurology outpatient settings with limited face-to-face time per patient need a short but specific memory test. The TMA-93 could fill the gap. The test runs with a ceiling effect in HCs and is highly discriminative for diagnosing patients with aMCI or mild dementia [16]. However, a floor effect should be expected in patients with moderate dementia and may could be already present in some patients with mild dementia, here represented by the outliers for whom the administration of the test took longer than 15 min. The target of the TMA-93 are mainly patients with memory complaints and no functional impairment when total scores on MMSE or MoCA are around the cutoffs and are not conclusive [24]. Studies comparing the diagnostic accuracy and feasibility of the TMA-93 against screeners, as MMSE or MoCA, in settings with limited face-to-face time per patient are needed.

The samples here tested were composed of a relatively high percentage of low-educated participants. Lack of education remains a limitation in many elderly Spanish people since they had limited primary school access in the aftermath of the Spanish Civil War (1936-1939). Although the situation has significantly improved in recent years, 59% of the population over 65 years of age in Spain did not complete primary studies [45]. Low-education is also a limitation for people in many developing countries in the world. In most developed countries, multicultural individuals with a different primary language, not proficient in the host country one, also have this limitation. The neuropsychological examination must comply with this handicap. Here, the TMA-93 was again demonstrated feasible to be administered to low-educated individuals. In fact, there were no significant differences in administration time by educational attainment. Despite this feasibility, the TMA-93 total score should be expected lower in low-educated individuals. Feasibility does not mean that the test is free of educational bias. Associative learning is also trained and acquired at school and, accordingly, normative studies show lower TMA-93 total score in less educated groups [24, 38].

In addition to optimal diagnostic accuracy previously reported for the TMA-93, the good reliability and feasibility here demonstrated encourages the completion of the test´s development. The next steps will be phase II and III validation studies, including AD biomarkers and comparing the diagnostic accuracy of the test with that of the standard mem-

ory instruments on samples organized by educational attainment.

Conclusion

In summary, our findings of good reliability (internal consistency and inter-rater and test-retest reliability) and feasibility (task-tolerability, short administration time, and simplicity of administration and scoring after some training) make the TMA-93 a brief relational binding memory test suitable to be administered to patients with memory complaints, particularly in settings with limited face-to-face time per patient and low-educated population.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

REFERENCES

[1] Parra MA, Abrahams S, Fabi K, Logie R, Luzzi S, Della Sala S (2009) Short-term memory binding deficits in Alzheimer's disease. *Brain* **132**, 1057-1066.

[2] Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* **12**, 97-136.

[3] Zimmer H, Mecklinger A, Lindenberger U (2006) Levels of binding: Types, mechanisms, and functions of binding in remembering. *Handbook of Binding and Memory, perspective from cognitive neuroscience*, Zimmer HD, Mecklinger A, Lindenberger U, eds. Oxford University Press, New York, pp. 3-22.

[4] Pavisic IM, Suarez-Gonzalez A, Pertzov Y (2020) Translating visual short-term memory binding tasks to clinical practice: From theory to practice. *Front Neurol* **11**, 458.

[5] Mayes A, Montaldi D, Migo E (2007) Associative memory and the medial temporal lobes. *Trends Cogn Sci* **11**, 126-135.

[6] Parra MA (2014) Overcoming barriers in cognitive assessment of Alzheimer's disease. *Dement Neuropsychol* **8**, 95-98.

[7] Parra MA, Abrahams S, Logie RH, Della Sala S (2010) Visual short-term memory binding in Alzheimer's disease and depression. *J Neurol* **257**, 1160-1169.

[8] Swainson R, Hodges JR, Galton CJ, Semple J, Michael A, Dunn BD, Iddon JL, Robbins TW, Sahakian BJ (2001) Early detection and differential diagnosis of Alzheimer's disease

and depression with neuropsychological tasks. *Dement Geriatr Cogn Disord* **12**, 265-280.

[9] Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ (2002) Paired associate performance in the early detection of DAT. *J Int Neuropsychol Soc* **8**, 58-71.

[10] Parra MA, Saarimäki H, Bastin ME, Londoño AC, Pettit L, Lopera F, Della Sala S, Abrahams S (2015) Memory binding and white matter integrity in familial Alzheimer's disease. *Brain* **138**, 1355-1369.

[11] Rentz DM, Locascio JJ, Becker JA, Moran EK, Enq E, Buckner RL, Sperling RA, Johnson KA (2010) Cognition, reserve, and amyloid deposition in normal aging. *Ann Neurol* **67**, 353-364.

[12] Wechsler D (1997) *Wechsler Memory Scale-Third Edition*. The Psychological Corporation, San Antonio, TX.

[13] Buschke H (2014) The rationale of the Memory Binding Test. *Dementia and Memory*, Nilsson LG, Ohta N, eds. Psychology Press, New York, pp. 55-71.

[14] Rentz DM, Amariglio RE, Becker JA, Frey M, Olson LE, Frishe K, Carmasin J, Maye JE, Johnson KA, Sperling RA (2011) Face-name associative memory performance is related to amyloid burden in normal elderly. *Neuropsychologia* **49**, 2776-2783.

[15] Maillet D, Narme P, Amieva H, Matharan F, Bailon O, Le Clésiau H, Belin C (2017) The TMA-93: A new memory test for Alzheimer's disease in illiterate and less educated people. *Am J Alzheimers Dis Other Dement* **32**, 461-467.

[16] Rodrigo-Herrero S, Carnero-Pardo C, Méndez-Barrio, C, De Miguel-Tristancho M, Graciani-Cantisán E, Sánchez-Arjona MB, Maillet D, Jiménez-Hernández MD, Franco-Macías E (2019) TMA-93 for diagnosing amnestic mild cognitive impairment: A comparison with the Free and Cued Selective Reminding Test. *Am J Alzheimers Dis Other Dement* **34**, 322-328.

[17] Welsh KA, Butters N, Hughes J, Mohs R, Heyman AD (1991) Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Arch Neurol* **48**, 278-281.

[18] Moradi E, Hallikainen I, Hänninen T, Tohka J; Alzheimer's Disease Neuroimaging Initiative (2017) Rey's Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer's disease. *Neuroimage Clin* **13**, 415-427.

[19] Costa A, Bak T, Caffarra P, Caltagirone C, Ceccaldi M, Collete F, Crutch S, Della Sala S, Démonet JF, Dubois B, Duzel E, Nestor P, Papageorgiou SG, Salmon E, Sikkes S, Tiraboschi P, Van der Flier W, Visser PJ, Cappa SF (2017) The need for harmonization and innovation of neuropsychological assessment in neurodegenerative dementias in Europe: Consensus document of the Joint Program for Neurodegenerative Diseases Working Group. *Alzheimers Res Ther* **9**, 27.

[20] Carnero Pardo C, Sáez-Zea C, Montiel Navarro L, Del Sazo P, Feria Vilar I, Pérez Navarro MJ, Ruiz Jiménez J, Vilchez Carrillo R, Montoro Ríos MT (2007) Diagnostic accuracy of the Phototest for cognitive impairment and dementia. *Neurologia* **22**, 860-869.

[21] Barbeau E, Didic M, Tramoni E, Felician O, Joubert S, Sontheimer A, Ceccaldi M, Poncet M (2004) Evaluation of visual recognition memory in MCI patients. *Neurology* **62**, 1317-1322.

[22] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's

disease: Recommendations from the National Institute on Aging and Alzheimer's Association workgroup. *Alzheimers Dement* **7**, 270-279.

[23] Teunisse S, Derix MM (1997) The interview for deterioration in daily living activities in dementia: Agreement between primary and secondary caregivers. *Int Psycogeriatr* **9**, 155-162.

[24] Rodrigo-Herrero S, Sánchez-Benavides G, Ainz-Gómez L, Luque-Tirado A, Graciani-Cantisán E, Bernal Sánchez-Arjona M, Maillet D, Jiménez-Hernández MD, Franco-Macías E (2020) Norms for testing visual binding using the Memory Associative Test (TMA-93) in older educationally-diverse adults. *J Alzheimers Dis* **34**, 322-328.

[25] Carnero Pardo C, Carrera Muñoz I, Triguero Cueva L, López Alcalde S, Vílchez Carrillo R (2018) Normative data for the Fototest from neurological patients with no cognitive impairment. *Neurologia*, doi: 10.1016/j.nrl.2018.03.001

[26] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JI, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263-269.

[27] Reisberg B, Ferris SH, de León MJ, Crook T (1982) The global deterioration scale for assessment of primary degenerative dementia. *Am J Psychiatry* **139**, 1136-1139.

[28] Carnero Pardo C, De la Vega Cotarelo R, López Alcalde S, Martos Aparicio C, Vílchez Carrillo R, Mora-Gavilán E, Galvin JE (2013) Assessing the diagnostic accuracy (DA) of the Spanish version of the informant-based AD8 questionnaire. *Neurologia* **28**, 88-94.

[29] Grober E, Sanders AE, Hall C, Lipton RB (2010) Free and Cued Selective Reminding Test identifies very mild dementia in primary care. *Alzheimer Dis Assoc Disord* **24**, 284-290.

[30] Golden CJ (1978) *Stroop Color and Word Test: A Manual for Clinical and Experimental Uses*. Stoelting Co., Chicago, IL.

[31] Cano S, Posner H, Moline M, Hurt S, Swartz J, Hsu T, Hobart J (2010) The ADAS-Cog in Alzheimer´s disease clinical trials: Psychometric evaluation of the sum and its parts. *J Neurol Neurosurg Psychiatry* **81**, 1363-1638.

[32] Serrano C, Allegri RF, Drake M, Butman J, Harris P, Nagle C, Ranalli C (2001). A shortened form of the Spanish Boston naming test: A useful tool for the diagnosis of Alzheimer's disease. *Rev Neurol* **33**, 624-627.

[33] Boyd CD, Tierney M, Wassermann EM, Spina S, Oblak AL, Ghetti B, Grafman J, Huey E (2014) Visuoperception test predicts pathologic diagnosis of Alzheimer disease in corticobasal syndrome. *Neurology* **83**, 510-519.

[34] Martínez de la Iglesia J, Onís Vilches MC, Dueñas Herrero R, Aguado Taberné C, Albert Colomer C, Luque Luque R (2005). Abbreviating the brief. Approach to ultra-short versions of the Yesavage questionnaire for the depression. *Aten Primaria* **35**, 14-21.

[35] Nunnally JC, Bernstein IH (1994) Psychometric theory. 3rd ed, McGraw-Hill, New York.

[36] Portney LG, Watkins MP (2015). Foundations of clinical research: Applications to practice. 3rd edition, Davis Company, Philadelphia.

[37] Gliem JA, Gliem RR (2003) Calculating, interpreting, and reporting Cronbach's alpha reliability for Likert-type scales. *Midwest Research to Practice Conference in Adult, Continuing, and Community Education*, Columbus, pp. 82-88.

[38] Dessi F, Maillet D, Metivet E, Michault A, Le Clésiau H, Ergis AM, Belin C (2009) Assessment of episodic memory in illiterate elderly. *Psychol Neuropsychiatr Vieil* **7**, 287-296.

[39] Welch S, Comer J (1988) *Quantitative Methods for Public Administration: Techniques and Applications*. Books/Cole Publishing Co.

[40] Clerici F, Ghiretti R, Di Pucchio A, Pomati S, Cucumo V, Marcone A, Vanacore N, Mariani C, Cappa SF (2017) Construct validity of the Free and Cued Selective Reminding Test in older adults with memory complaints. *J Neuropsychol* **11**, 238-251.

[41] O'Connor DW, Pollit PA, Hyde JB, Fellows JL, Miller ND, Brook CP, Reiss BB (1989) The reliability and validity of the Mini-Mental State in a British community survey. *J Psychiatr Res* **23**, 87-96.

[42] Gramunt N, Sánchez-Benavides G, Buschke H, Lipton RB, Masramon X, Gispert JD, Peña-Casanova J, Fauria K, Molinuevo JL (2016) Psychometric properties of the Memory Binding Test: Test-retest reliability and convergent validity. *J Alzheimers Dis* **50**, 999-1010.

[43] Guttman L (1945) A basis for analyzing test-retest reliability. *Psychometrika* **10**, 255-282.

[44] Zulfiqar AA, De Longuemarre AT (2019) Codex and MMSE: What to choose? *Geriatr Psychol Neuropsychiatr Vieil* **17**, 279-289.

[45] De la Fuente A, Doménech R (2016) El nivel educativo de la población en España y sus regiones: 1960-2011. *Investigaciones Regionales* **34**, 73-94.