Memory Binding Test Distinguishes Amnestic Mild Cognitive Impairment and Dementia from Cognitively Normal Elderly

Herman Buschke[1,2,†], Wenzhu B. Mowrey[2,3,†*], Wendy S. Ramratan[1,2], Molly E. Zimmerman[1,2], David A. Loewenstein[4], Mindy J. Katz[1,2], Richard B. Lipton[1,2]

[1]*Department of Neurology, Albert Einstein College of Medicine, Bronx, NY 10461, USA*
[2]*The Einstein Aging Study, Albert Einstein College of Medicine, Bronx, NY 10461, USA*
[3]*Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA*
[4]*Department of Psychiatry and Behavioral Sciences, University of Miami, Miami, FL 33136, USA*

*Corresponding author at: Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA. Tel.: +1-718-430-2765; fax: +1-718-430-2608.
E-mail address: wenzhu.mowrey@einstein.yu.edu (W. B. Mowrey).

Accepted 6 September 2016

Abstract

Objective: We aimed to assess reliability and cross-sectional discriminative validity of the Memory Binding Test (MBT) to distinguish persons with amnestic cognitive impairment (aMCI) and dementia from cognitively normal elderly controls.
Method: The MBT was administered to 20 participants with dementia, 31 with aMCI and 246 controls, who received the first administration of the MBT from May 2003 to December 2007, as a substudy of the community-based Einstein Aging Study (age range: 70+). The optimal index resulted from comparing the partial area under the receiver operating characteristic curves (ROC AUC) of four major MBT indices for specificities ≥ 0.70. Optimal cut-score of the optimal index was selected by maximizing the sum of sensitivity and specificity. Age and education effects were assessed using stratified cut-scores and adjusted logistic regression. Reliability was computed as intraclass correlation between scores at baseline and 1-year follow-up for participants who remained cognitively normal.
Results: Total number of Items recalled in the Paired condition (TIP) was elected the optimal index. TIP cut-score was ≤ 22 for differentiating aMCI alone (sensitivity = 0.74, specificity = 0.73) and aMCI and dementia combined (sensitivity = 0.84, specificity = 0.73) from controls. It was ≤ 17 for differentiating dementia from aMCI and controls (sensitivity = 0.95, specificity = 0.87). Age and education adjustments did not materially improve discriminative validity. The reliability of TIP was 0.77.
Conclusions: MBT achieved moderate to good reliability. TIP had superior cross-sectional discriminative validity than the other MBT indices. We recommend using the empirical cut-score of TIP ≤ 22 for discriminating aMCI and dementia and ≤ 17 for discriminating dementia alone.

Keywords: Dementia; Alzheimer's disease; Mild cognitive impairment; Learning and memory; Elderly/geriatrics/aging

Introduction

Improving the early detection of Alzheimer's disease (AD) has emerged as an urgent public health priority (Dubois et al., 2010; Jack et al., 2012; Sperling et al., 2011). There is a pressing need to develop simple and reliable cognitive screening measures that can detect subtle cognitive variations in the prodromal stages since current standard tests used in the diagnoses of mild cognitive impairment (MCI) and dementia were not designed to detect such subtle cognitive variations (Rentz et al., 2013). One approach to improving cognitive tests is to target cognitive processes thought to be affected early in the prodromal stages of AD. One such process, memory binding, refers to the ability to retrieve independent information units as part of a

---

† These authors contributed equally to the manuscript.

complex unit (Amariglio et al., 2012; Buschke, 2014; Parra et al., 2009, 2010; Villeneuve & Belleville, 2012). Deficits on episodic binding of shape and color, probed by the Short-Term Memory Binding test (STMB), are present in AD, familial AD and asymptomatic carriers of the E280A single presenilin-1 mutation (Parra et al., 2009, 2010). Associative memory of face and name, detected by the Face–Name Associative Memory Exam (FNAME), is related to amyloid burden in cognitively normal elderly (Rentz et al., 2011). These findings suggest that memory binding may serve as a behavioral marker for early detection.

Buschke and colleagues developed the Memory Binding Test (MBT) (Buschke, 2014), previously known as Memory Capacity Test (Papp et al., 2015; Rentz et al., 2010, 2013), to probe semantic binding in hopes that it might detect subtle cognitive variations in the pre-MCI stage. Like the Free and Cued Selective Reminding Test (FCSRT; Buschke, 1984; Grober & Buschke, 1987), the MBT employs controlled learning and cued recall to manage encoding specificity and elicit maximum retrieval. While the FCSRT does not involve learning of paired items, the MBT requires participants to learn two word items (e.g., tulip and carnation) associated with a single shared semantic category cue (e.g., flower). MBT performance is associated with the prodromal stages of AD as defined by biomarkers (Jack et al., 2012; Papp et al., 2015; Rentz et al., 2010). In the current study, we aimed to assess the reliability of the MBT and its cross-sectional discriminative validity for distinguishing persons with amnestic MCI (aMCI) and dementia from persons without memory impairment, that is the reference group, in a community-based sample. We employed two analysis strategies. In the primary analyses, cognitively normal controls alone form the reference group of persons without memory impairment; in the secondary analyses, in order to provide evidence for using the MBT as a more practical screening test without the administration of a full neuropsychological battery, controls and persons with non-amnestic MCI (naMCI) were combined to form a second reference group, which more closely replicates settings where individuals do not receive a full neuropsychological battery including assessments of non-memory domains to differentiate persons with naMCI from cognitively normal.

Methods

Participants

This present study included 329 participants who received the first administration of the basic version, that is Version 1 of the MBT (Gramunt et al., 2015; Gramunt, Sanchez-Benavides, Buschke, Dieguez-Vide, et al., 2016; Gramunt, Sanchez-Benavides, Buschke, Lipton, et al., 2016; Papp et al., 2015; Rentz et al., 2010), from May 2003 to December 2007 as a substudy of the Einstein Aging Study (EAS). The EAS is a community-based study that assesses a systematically recruited sample of older adults residing in Bronx, NY (Katz et al., 2012). EAS inclusion criteria were age 70+, English speaking, community-residing (noninstitutionalized) and ambulatory. Individuals were excluded from EAS if they had sensorimotor impairment too severe to perform the neuropsychological tests. The study protocols were approved by the local Institutional Review Board and informed consent was obtained from each participant. This subset of 329 participants receiving Version 1 of the MBT constituted 39% of all the 848 participants evaluated by the parent study, EAS, during the study time period from May 2003 to December 2007. Participants receiving Version 1 were similar in terms of demographic characteristics and cognitive performance to the remaining participants evaluated by EAS during the same time period, which included 301 participants (35%) who received Version 2 of the MBT, a significantly different procedure from Version 1 (Buschke, 2014), and the remaining 218 participants (26%) who did not receive any version of the MBT (see Supplementary material online, Table S1).

EAS Assessments

Each participant received the standard EAS assessment battery including demographic information, medical history, health status, a standard neurologic exam, and a neuropsychological test battery at baseline and at annual visits (Katz et al., 2012). For this present study, baseline was defined as the first administration of the MBT. Education was measured by years of education and by the Wide Range Achievement Test-Third Edition (WRAT-3) reading subtest grade score (Wilkinson, 1993). The WRAT-3 reading subtest measures the ability to recognize and pronounce words with increasing complexity and unfamiliarity and produces a grade score representing the number of years of education equivalent to the achieved reading level. Functional status was assessed by clinical evaluation, the clinical history forms (Cognitive impairment/dementia forms: C1 Alt for the study participant and C2-Alt for the informant about the participant) which are included in the Clinical and Neuropsychology Assessment of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) (Morris et al., 1989), and the instrumental activities of daily living (IADL) Scale, a subscale of the Lawton Brody Scale (Lawton & Brody, 1969).

Depression was assessed by the 15-item Geriatric Depression Scale (GDS, range: 0−15) with a score of ≥5 indicating clinical depression (Julian et al., 2009).

Global cognitive function was assessed by the Blessed Information-Memory-Concentration test (BIMC, range: 0−33) where a higher score indicates poorer performance (Blessed, Tomlinson, & Roth, 1968; Katzman et al., 1988). Episodic memory was measured by the FCSRT free recall score (FCSRT-FR, range: 0−48) (Buschke, 1984; Grober & Buschke, 1987) and Logical Memory I subtest of the Wechsler Memory Scale—Revised (LM-I, range: 0−50) (Wechsler, 1987). A higher score indicates better performance on both tests. Subjective memory impairment was assessed using self- and informant-reported information on the clinical history form included in the CERAD (Morris et al., 1989), the Informant Questionnaire on Cognitive Decline in the Elderly (Jorm & Jacomb, 1989), or the EAS Health Self-Assessment Questionnaire (Rabin et al., 2015). Language was assessed by the Category Fluency task (animals, vegetables, and fruits) (Lezak, Howieson, Bigler, & Tranel, 2012) and the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983). Attention was assessed by the Trail Making Test part A (Reitan, 1958) and the Digit Span subtest of the Wechsler Adult Intelligence Scale-III (Wechsler, 1997). Visuospatial construction was assessed by the Block Design subtest and the Digit Symbol subtest, both from Wechsler Adult Intelligence Scale-III (Wechsler, 1997). Executive function was assessed by the Trail Making Test part B (Reitan, 1958) and the Letter Fluency "FAS" task (Benton & Hamsher, 1989).

Diagnostic Criteria

Clinical diagnoses were assigned at consensus case conferences attended by a neurologist and a neuropsychologist. A dementia diagnosis was based on the criteria from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (American Psychiatric Association, 1994), which required impairment in memory and impairment in at least one other non-memory cognitive domain, and significant functional decline. Memory impairment was defined as FCSRT-FR ≤24 or 1.5 standard deviation (SD) or greater below the age-adjusted mean on LM-I. The optimal cut-scores for LM-I vary with age while optimal cut-scores for the FCSRT-FR have been shown not to vary with age. For example, in the Mayo's Older Americans Normative Studies (Steinberg, Bieliauskas, Smith, & Ivnik, 2005) optimal LM-I cut-scores were age-dependent. In the Baltimore Longitudinal Study of Aging and in the EAS a FCSRT-FR cut-score (≤24) did not need age adjustment (Grober & Kawas, 1997; Grober, Lipton, Hall, & Crystal, 2000). Impairment in the non-memory domains was defined as 1.5 SD or greater below the age-adjusted mean using tests listed above. Functional decline was determined at case conferences based on clinical evaluation, self- or informant-reports on the CERAD, and impaired score on the IADL Scale (Lawton & Brody, 1969). Dementia subtypes were assigned according to established criteria. Common subtype assignments included but were not restricted to: Probable/possible AD based on the clinical criteria established by the National Institute of Neurological and Communication Disorders and Stroke and the Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA criteria) (McKhann et al., 1984), probable/possible vascular dementia (California criteria) (Chui et al., 1992), and frontotemporal dementia (FTD; Neary criteria) (McKhann et al., 2001; Neary et al., 1998). When more than one etiology was clinically evident, multiple diagnoses were assigned.

An aMCI diagnosis was assigned based on revised Peterson criteria (Artero, Petersen, Touchon, & Ritchie, 2006) and required objective memory impairment defined as above, subjective memory impairment reported by self or informants, no evidence of functional decline, no diagnosis of dementia, with or without impairment in other non-memory cognitive domains. A naMCI diagnosis was assigned to participants who did not meet criteria for dementia, had no functional decline, did not meet the criteria for memory impairment but had impairment (1.5 SD or greater below the age-adjusted mean) in one or more non-memory cognitive domains. Persons meeting criteria for aMCI with impairment in one or more non-memory cognitive domains were classified as aMCI in this study. Individuals not meeting any of these criteria were diagnosed as cognitively normal.

Memory Binding Test

The MBT consists of 16 category cues (e.g., flower) and two word lists containing 16 word items in each list (e.g., tulip in the first list, carnation in the second list), where the categories are presented in the same order in both lists (Buschke, 2014). In this basic version, that is Version 1, both lists are learned and recalled by controlled learning and cued recall, followed by a last step, the paired recall condition from both lists (Buschke, 2014; Gramunt et al., 2015; Gramunt, Sanchez-Benavides, Buschke, Dieguez-Vide, et al., 2016; Gramunt, Sanchez-Benavides, Buschke, Lipton, et al., 2016; Papp et al., 2015; Rentz et al., 2010). The test procedure is the following. In the first step, the participant was instructed to learn the 16 word items from the first list, presented on 4 cards with 4 words on each card, 1 card at a time. For each item, after the examiner stated a

category cue (e.g., flower), the participant was asked to verbally identify the correct item from the card associated with the cue (e.g., tulip) within 5 s to ensure controlled learning and encoding specificity. In the second step, the participant was instructed to recall the items from the first list after cue presentation. Five seconds were allowed for each cue and errors were not corrected. In the third step, the participant was instructed to learn 16 new items from the second list, as similarly done for the first list. In the fourth step, the participant was asked to recall items from the second list, as similarly done for the first list. In the fifth step, the paired recall condition, for each cue the participant was asked to recall both items from both lists in any order within 10 s. The total procedure takes about 6 min.

The MBT generates four major indices (see Supplementary material online, Table S2) (Gramunt et al., 2015; Gramunt, Sanchez-Benavides, Buschke, Dieguez-Vide, et al., 2016; Gramunt, Sanchez-Benavides, Buschke, Lipton, et al., 2016; Papp et al., 2015): The number of items cued recalled from List 1 (CR-L1, range 0−16) obtained in the second step, the number of items cued recalled from List 2 (CR-L2, range 0−16) obtained in the fourth step, the number of Pairs cued recalled In the Paired condition (PIP, range 0−16) and the Total number of Items cued recalled in the Paired condition (TIP, range 0−32) obtained in the fifth step. TIP equals to "PIP×2 + the number of single items recalled in the paired condition". PIP and TIP are measures on memory binding. "CR-L1−CR-L2" was used to assess the effects of semantic interference (Loewenstein et al., 2003).

Statistical Methods

Demographic characteristics and neuropsychological performance were summarized for each diagnostic group at baseline. Kruskal–Wallis test was used to compare continuous variables among ≥3 groups and Wilcoxon–Mann–Whitney test was used to compare between two groups. Pearson's chi-square or Fisher's exact tests were used to compare categorical variables. Test-retest reliability was computed as intraclass correlation between scores at baseline and 1-year follow-up: A) for the participants with both visits, B) for the participants who were cognitively normal at both visits, and C) for those whose global cognitive function remained stable as measured by the BIMC (absolute change ≤1), respectively. Discriminative validity of the MBT was assessed by the following comparisons at baseline: A) aMCI versus controls, B) aMCI and dementia versus controls, and C) Dementia versus controls and aMCI. Receiver operating characteristic (ROC) curves were plotted for all comparisons using each MBT index. The optimal MBT index was chosen by comparing the partial area under the curve (AUC) for specificities ≥0.70 (Robin et al., 2011). The optimal cut-score of the optimal MBT index was chosen to maximize the sum of sensitivity and specificity although we acknowledge that the optimal cut-score may differ in other scenarios since the tradeoffs between sensitivity and specificity depend on the goals of the application (see Discussion). Two approaches were applied to assess the effects of age and education on discriminative validity. One approach was to compute age and education stratified cut-scores defined as "mean−1.5×SD" in each age and education stratum of the reference group. These scores were applied to each individual to compute sensitivity and specificity, which were compared to operating characteristics results without age and education adjustment using McNemar's tests. The other approach was to adjust for age and education in logistic regression models, where the ROC AUC was compared to results obtained from logistic regression models without adjustment. In the secondary analysis strategy, all analyses were repeated by including the naMCI group, where controls and naMCI were combined to form the reference group for the following comparisons: A') aMCI versus controls and naMCI, B') aMCI and dementia versus controls and naMCI, and C') Dementia versus controls, naMCI and aMCI.

Results

Among the 329 older adults who received the first administration of the MBT, 20 participants were diagnosed with dementia, 31 with aMCI, 32 with naMCI, and the remaining 246 were cognitively normal at baseline (Table 1). Dementia subtypes included 5 cases diagnosed with both AD and vascular dementia, 10 cases with AD only (15 AD cases in total by including the 5 cases diagnosed with both AD and vascular dementia), 4 cases with vascular dementia only (9 vascular dementia cases in total by including the 5 cases diagnosed with both AD and vascular dementia), and 1 case of FTD. Four diagnostic groups were comparable with respect to gender and education. Persons with aMCI were older than controls; persons with dementia were older than controls and persons with naMCI. Race was borderline significant (*p*=.053), probably due to the higher proportion of blacks in the naMCI group. The aMCI and dementia groups were more depressed compared to controls. Global cognitive function worsened from controls, to naMCI and aMCI, and then to dementia, indicated by higher BIMC scores. Compared to controls, persons with naMCI were very similar except for a higher percentage of blacks, worse global function and lower but still within normal range memory scores on FCSRT-FR and LM-I.

Table 1. Demographic characteristics and neuropsychological test performance by diagnostic group at baseline[a]

	Controls (n = 246)	naMCI (n = 32)	aMCI (n = 31)	Dementia (n = 20)	4-group comparison p value
Age, mean (SD), years	79.7 (4.6)	80.1 (5.0)	82.3 (4.6)[b]	83.6 (4.9)[c,d]	0.0005
Female, % (n)	59.4 (146)	62.5 (20)	51.6 (16)	75.0 (15)	0.41
Race, % (n)					
White	67.9 (167)	40.6 (13)[e]	71.0 (22)[f]	60.0 (12)	
Black	27.6 (68)	53.1 (17)[e]	29.0 (9)[f]	40.0 (8)	0.053
Other	4.5 (11)	6.3 (2)[e]	0 (0)[f]	0 (0)	
Years of education	13.8 (3.4)	13.1 (3.5)	12.4 (3.5)	12.9 (4.0)	0.22
WRAT-3	11.5 (2.3)	10.7 (2.7)	11.1 (2.9)	11.0 (2.7)	0.30
GDS score	2.1 (2.2)	3.2 (3.3)	3.5 (2.8)[b]	4.9 (3.9)[c]	<0.0001
BIMC	1.9 (1.9)	2.7 (2.0)[e]	3.8 (2.8)[b]	8.7 (3.5)[c,d,g]	<0.0001
FCSRT-FR	32.8 (5.6)	30.3 (4.6)[e]	19.4 (4.5)[b,f]	13.7 (6.8)[c,d,g]	<0.0001
LM-I	22.1 (7.1)	19.0 (5.0)[e]	16.6 (6.2)[b]	10.8 (5.7)[c,d,g]	<0.0001
MBT CR-L1	14.4 (1.7)	14.2 (1.9)	12.2 (2.6)[b,f]	8.7 (4.0)[c,d,g]	<0.0001
MBT CR-L2	11.6 (2.9)	10.8 (3.4)	9.2 (3.2)[b]	4.8 (3.0)[c,d,g]	<0.0001
MBT PIP	10.7 (3.2)	10.0 (3.4)	7.5 (3.2)[b,f]	3.1 (2.5)[c,d,g]	<0.0001
MBT TIP	24.9 (5.0)	23.6 (5.3)	18.5 (6.1)[b,f]	10.0 (5.0)[c,d,g]	<0.0001

Notes: naMCI, non-amnestic mild cognitive impairment; aMCI, amnestic mild cognitive impairment; WRAT-3, Wide Range Achievement Test-Third Edition reading subtest grade score; GDS, Geriatric Depression Scale; BIMC, Blessed Information-Memory-Concentration test; FCSRT-FR, Free and Cued Selective Reminding Test free recall score; LM-I, Logical Memory I; MBT, Memory Binding Test; CR-L1, Number of items cued recalled from List 1 on the MBT; CR-L2, Number of items cued recalled from List 2 on the MBT; PIP, Number of Pairs cued recalled In the Paired condition on the MBT; TIP, Total number of Items cued recalled in the Paired condition on the MBT.

[a]For continuous variables, Kruskal–Wallis tests were used for comparing among three or more groups and Wilcoxon–Mann–Whitney tests for comparing between two groups. For categorical variables, Pearson's chi-square or Fisher's exact tests were used. Significance level was defined as $p < .05$.
[b]Significant difference between controls and aMCI.
[c]Significant difference between controls and dementia.
[d]Significant difference between naMCI and dementia.
[e]Significant difference between controls and naMCI.
[f]Significant difference between naMCI and aMCI.
[g]Significant difference between aMCI and dementia.

Primary Results—Reference Group Included Cognitively Normal Controls Alone

The primary analysis excluded persons with naMCI and used a reference group that included only cognitively normal controls. The MBT achieved moderate to good test-retest reliability (Table 2), especially for the MBT index TIP. The reliability of TIP was 0.85 for the 200 participants tested at both baseline and 1-year. When we restricted the sample to alleviate the effects of the real cognitive change on reliability, the reliability of TIP was 0.77 for 156 participants who remained cognitively normal for 1 year and 0.83 for the 137 participants with sustained global cognitive function defined by absolute change in BIMC ≤ 1. These reliability values of the MBT were higher compared to those of FSCRT-FR and LM-I.

ROC curves were plotted in Fig. 1. Tabular results of sensitivity and specificity on all indices are listed in Table 3 for TIP and Supplementary Table S3 for other indices. Comparisons of partial AUC for specificities ≥ 0.70 were summarized (Table 4). TIP achieved significantly higher partial AUC than CR-L2 ($p = .01$) in distinguishing aMCI from controls, than CR-L1 ($p = .01$), CR-L2 ($p = .003$) and PIP ($p = .03$) in distinguishing aMCI and dementia from controls, than CR-L1 ($p = .02$) and CR-L2 ($p = .04$) in distinguishing dementia from controls and aMCI. PIP was only superior compared to one other index CR-L2 in one comparison distinguishing aMCI and dementia from controls ($p = .02$), so we chose TIP to be the optimal index.

Based on the rule of maximizing the sum of sensitivity and specificity (Table 3), the optimal cut-score for TIP was ≤22 to distinguish aMCI from controls (sensitivity = 0.74, specificity = 0.73), and the optimal cut-score was ≤17 to distinguish dementia from controls and aMCI (sensitivity=0.95, specificity = 0.87). When the goal was to distinguish aMCI and dementia from controls, the same cut-score TIP ≤ 22 achieved a sensitivity of 0.84 and a specificity of 0.73. Even though the cut-score TIP ≤17 achieved a slightly higher sensitivity and specificity sum (1.581 vs. 1.575), we decided to retain the same cut-score ≤22 for distinguishing aMCI alone and aMCI and dementia together because having this same cut-score simplifies clinical applications and delivers higher sensitivity for screening purposes.

The discriminative validity of semantic interference (Loewenstein et al., 2003) measure "CR-L1−CR-L2" was not as good as any of the four major indices (see Fig. 1 and Supplementary material online, Table S3).

Table 2. MBT test-retest reliability over 1 year for all participants, stable controls, and persons with minimal BIMC change[a]

Cognitive measure	Entire reliability sample[b] (n = 200)	Stable controls[c] (n = 156)	BIMC absolute change ≤1[d] (n= 137)
MBT CR-L1	0.58	0.44	0.52
MBT CR-L2	0.68	0.62	0.67
MBT PIP	0.75	0.67	0.73
MBT TIP	0.85	0.77	0.83
FCSRT-FR	0.76	0.59	0.73
LM-I	0.77	0.73	0.78

Notes: MBT, Memory Binding Test; CR-L1, Number of items cued recalled from List 1 on the MBT; CR-L2, Number of items cued recalled from List 2 on the MBT; PIP, Number of Pairs cued recalled In the Paired condition on the MBT; TIP, Total number of Items cued recalled in the Paired condition on the MBT; BIMC, Blessed Information-Memory-Concentration test; FCSRT-FR, Free and Cued Selective Reminding Test free recall score; LM-I, Logical Memory I.
[a]Results excluded persons with naMCI at baseline. Supplementary Table S4 shows the corresponding reliability results that included persons with naMCI at baseline.
[b]There were 200 participants with data at both baseline and 1-year follow-up.
[c]Stable controls: Participants who were cognitively normal at both baseline and 1-year follow-up.
[d]Participants whose global cognitive function remained stable over 1 year. Global function was measured by the BIMC.
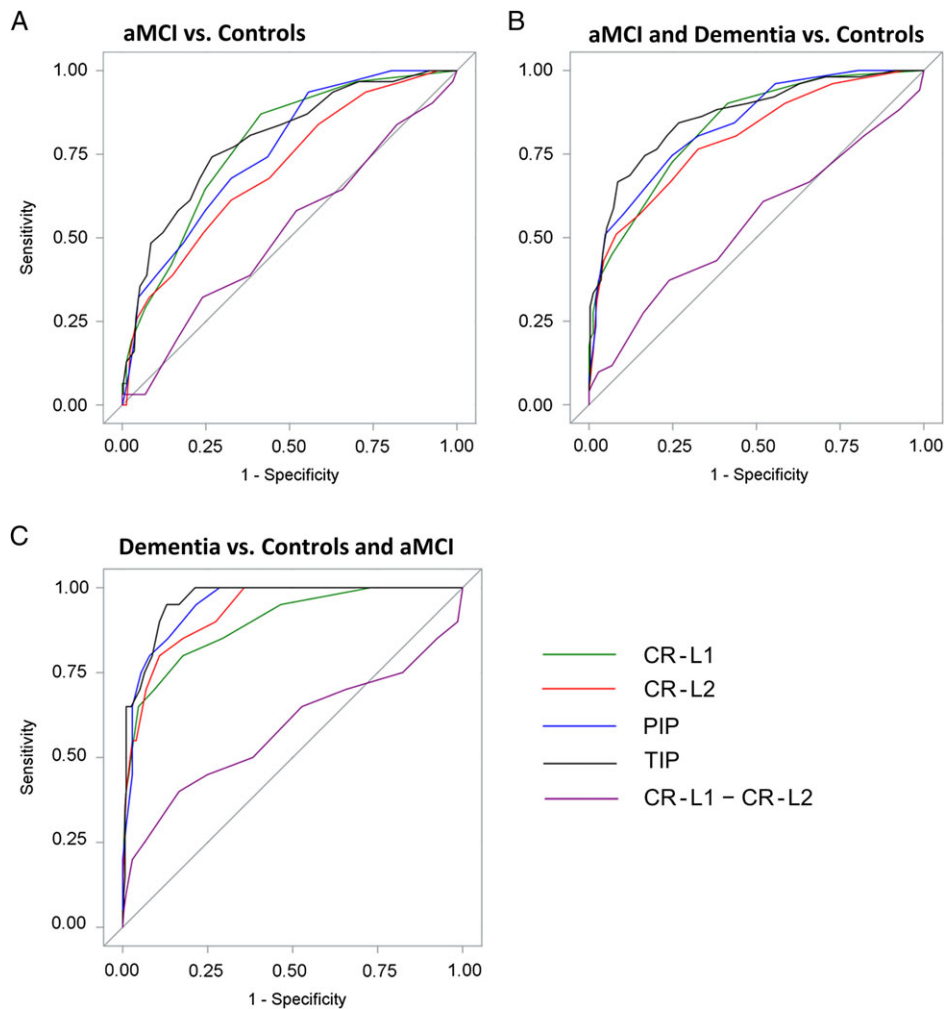

Fig. 1. Receiver operating characteristic curves for MBT indices CR-L1, CR-L2, PIP, TIP, and the semantic interference measure "CR-L1–CR-L2" (results excluded persons with naMCI; Supplementary Fig. S1 shows the corresponding results that included persons with naMCI).
MBT, Memory Binding Test; CR-L1, Number of items cued recalled om List 1 on the MBT; CR-L2, Number of items cued recalled from List 2 on the MBT; PIP, Number of Pairs cued recalled In the Paired condition on the MBT; TIP, Total number of Items cued recalled in the Paired condition on the MBT; aMCI, amnestic mild cognitive impairment.

Table 3. Sensitivity, specificity, and optimal cut-scores of the optimal MBT index TIP[a,b]

TIP cutoff	aMCI vs. controls		aMCI and dementia vs. controls		Dementia vs. controls and aMCI	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
≤13	0.16 (0.05,0.34)	0.96 (0.93,0.98)	0.37 (0.24,0.52)	0.96 (0.93,0.98)	0.70 (0.46,0.88)	0.95 (0.92,0.97)
≤14	0.26 (0.12,0.45)	0.96 (0.93,0.98)	0.45 (0.31,0.60)	0.96 (0.93,0.98)	0.75 (0.51,0.91)	0.94 (0.90,0.96)
≤15	0.35 (0.19,0.55)	0.95 (0.91,0.97)	0.53 (0.38,0.67)	0.95 (0.91,0.97)	0.80 (0.56,0.94)	0.91 (0.87,0.94)
≤16	0.39 (0.22,0.58)	0.93 (0.89,0.96)	0.59 (0.44,0.72)	0.93 (0.89,0.96)	0.90 (0.68,0.99)	0.89 (0.85,0.93)
≤17[c]	0.48 (0.30,0.67)	0.91 (0.87,0.95)	0.67 (0.52,0.79)	0.91 (0.87,0.95)	**0.95 (0.75,1.00)**	**0.87 (0.82,0.91)**
≤18	0.52 (0.33,0.70)	0.88 (0.83,0.92)	0.69 (0.54,0.81)	0.88 (0.83,0.92)	.95 (0.75,1.00)	0.83 (0.78,0.88)
≤19	0.58 (0.39,0.75)	0.83 (0.78,0.88)	0.75 (0.60,0.86)	0.83 (0.78,0.88)	1.00 (0.83,1.00)	0.79 (0.73,0.83)
≤20	0.61 (0.42,0.78)	0.80 (0.74,.85)	0.76 (0.63,0.87)	0.80 (0.74,0.85)	1.00 (0.83,1.00)	0.75 (0.70,0.80)
≤21	0.68 (0.49,0.83)	0.77 (0.71,0.82)	0.80 (0.67,0.90)	0.77 (0.71,0.82)	1.00 (0.83,1.00)	0.72 (0.66,0.77)
≤22[d]	**0.74 (0.55,0.88)**	**0.73 (0.67,0.79)**	**0.84 (0.71,0.93)**	**0.73 (0.67,0.79)**	1.00 (.83,1.00)	0.68 (0.62,0.73)
≤23	0.77 (0.59,0.90)	0.66 (0.60,0.72)	0.86 (0.74,0.94)	0.66 (0.60,0.72)	1.00 (0.83,1.00)	0.61 (0.55,0.67)
≤24	0.81 (0.63,0.93)	0.62 (0.55,0.68)	0.88 (0.76,0.96)	0.62 (0.55,0.68)	1.00 (0.83,1.00)	0.57 (0.51,0.63)
≤25	0.84 (0.66,0.95)	0.52 (0.46,0.59)	0.90 (0.79,0.97)	0.52 (0.46,0.59)	1.00 (0.83,1.00)	0.48 (0.42,0.54)
≤26	0.87 (0.70,0.96)	0.45 (0.38,0.51)	0.92 (0.81,0.98)	0.45 (0.38,0.51)	1.00 (0.83,1.00)	0.41 (0.35,0.47)
≤27	0.94 (0.79,0.99)	0.37 (0.31,0.43)	0.96 (0.87,1.00)	0.37 (0.31,0.43)	1.00 (0.83,1.00)	0.34 (0.28,0.39)
≤28	0.97 (0.83,1.00)	0.29 (0.24,0.35)	0.98 (0.90,1.00)	0.29 (0.24,0.35)	1.00 (0.83,1.00)	0.26 (0.21,0.32)

Notes: MBT, Memory Binding Test; TIP, The Total number of Items cued recalled in the Paired condition on the MBT; aMCI, amnestic mild cognitive impairment.
[a]Results excluded persons with naMCI. Supplementary Table S6 shows the corresponding results that included persons with naMCI.
[b]The optimal cut-scores were chosen to maximize the sum of sensitivity and specificity.
[c]The optimal cut-score for distinguishing dementia from controls and aMCI.
[d]The optimal cut-score for distinguishing aMCI or aMCI and dementia from controls.

Table 4. Comparisons of four MBT indices using the partial area under the receiver operating characteristic curve (ROC AUC) for specificities ≥.70[a]

Partial AUC pairwise comparison	aMCI vs. controls	aMCI and dementia vs. controls	Dementia vs.controls and aMCI
TIP vs. CR-L1	0.15 vs. 0.13 ($p = .09$)	**0.20 vs. 0.17 ($p = .01$)**	**0.26 vs. 0.22 ($p = .02$)**
TIP vs. CR-L2	**0.15 vs. 0.11 ($p = .01$)**	**0.20 vs. 0.16 ($p = .003$)**	**0.26 vs. 0.23 ($p = .04$)**
TIP vs. PIP	0.15 vs. 0.13 ($p = .06$)	**0.20 vs. 0.18 ($p = .03$)**	0.26 vs. 0.25 ($p = .12$)
PIP vs. CR-L1	0.13 vs. 0.13 ($p = .57$)	0.18 vs. 0.17 ($p = .25$)	0.25 vs. 0.22 ($p = .10$)
PIP vs. CR-L2	0.13 vs. 0.11 ($p = .10$)	**0.18 vs. 0.16 ($p = .02$)**	0.25 vs. 0.23 ($p = .06$)
CR-L2 vs. CR-L1	0.11 vs. 0.13 ($p = .46$)	0.16 vs. 0.17 ($p = .77$)	0.23 vs. 0.22 ($p = .69$)

Notes: MBT, Memory Binding Test; aMCI, amnestic mild cognitive impairment; CR-L1, Number of items cued recalled from List 1 on the MBT; CR-L2, Number of items cued recalled from List 2 on the MBT; PIP, Number of Pairs cued recalled In the Paired condition on the MBT; TIP, Total number of Items cued recalled in the Paired condition on the MBT.
[a]Results excluded persons with naMCI. Supplementary Table S5 shows the corresponding results that included persons with naMCI.

Age and education adjustment did not significantly improve discriminative validity. In the first approach, we generated age and education stratified cut-scores from the controls with strata defined for age groups 72−80 and 81−90 and years of education ≤12 and >12. The stratified cut-scores (TIP ≤20 for the stratum with higher education and younger age, ≤18 for higher education and older age, ≤15 for lower education and younger age, and ≤16 for lower education and older age) were applied to distinguish aMCI alone (sensitivity=0.48, specificity=0.90), to distinguish aMCI and dementia (sensitivity=0.63, specificity=0.90), and to distinguish dementia alone (sensitivity=0.89, specificity=0.85). At the sensitivities achieved by these stratified cut-scores, MBT TIP cut-scores obtained without adjustment achieved specificities at least as good. In the second approach, age and education were adjusted using logistic regression models. The AUC with adjustment was not significantly different from the AUC without adjustment for distinguishing aMCI (0.82 vs. 0.79, p=.18), for distinguishing aMCI and dementia (0.88 vs. 0.86, p=.13), and for distinguishing dementia alone (0.97 vs. 0.96, p=.35).

Secondary Results—Reference Group of Persons Without Memory Impairment Included Cognitively Normal Controls and Persons with naMCI

The secondary analyses included persons with naMCI to form a second reference group combining controls and persons with naMCI. These secondary results (see Supplementary material online, Fig. S1 and Tables S4–S7) were very similar to the

primary analysis results (Fig. 1, Tables 2–4, and Supplementary material online, Table S3). TIP was identified as the optimal index. TIP optimal cut-scores remained the same achieving the same sensitivities and slightly lower specificities (0.72 vs. 0.73; 0.86 vs. 0.87). The performance of the semantic interference measure was still not good (Loewenstein et al., 2003). Age and education adjustments did not improve the discriminative validity.

Discussion

Our study showed that the MBT achieved good test-retest reliability and good cross-sectional discriminative validity for distinguishing persons with aMCI and dementia from cognitively normal elderly controls. TIP, which measures the ability of participants to recall two separately learned items in response to a category cue, achieved better discriminative validity than the other MBT indices. Age and education adjustment did not materially change discriminative validity in this sample. We recommend using the empirical cut-score of MBT TIP ≤22 for distinguishing aMCI alone or aMCI and dementia combined from controls and ≤17 for distinguishing dementia from aMCI and controls. Semantic interference measures from the MBT did not appear to have good discriminative validity. These conclusions also held true when we used a second reference group combining controls and persons with naMCI.

TIP achieved moderate to good test-retest reliability between visits about a year apart. Reliability of behavioral tests may be better measured if test and retest are closer in time, for example 2−4 weeks. Since real cognitive change is less likely to occur in a shorter time period, test-retest reliability within 4 weeks may well be higher than the 1-year reliability of TIP (0.83) for the group of participants whose global cognitive function remained stable. The MBT was designed as a semantically mediated test through controlled learning, so the discriminative validity did not benefit from age and education adjustment. The remarkable similarities between the primary and secondary analyses results indicate that the performance of the MBT is robust and can be used to distinguish persons with memory impairment (aMCI, dementia) from persons without memory impairment (controls, naMCI) in settings where individuals do not receive assessments in non-memory domains to differentiate persons with naMCI from cognitively normal. The secondary analyses results provided evidence for using the MBT as a more practical screening test without the administration of a full neuropsychological battery.

Both PIP and TIP were indices from the paired condition. PIP (range: 0–16) counts the number of pairs recalled where the participant had to recall both paired word items correctly. TIP (range: 0–32) indexes the number of word items recalled in the paired condition. By definition, TIP equals to "PIP × 2 + number of single items recalled". TIP clearly provided better discriminative validity than PIP. We do not know if PIP or TIP provides the best measure of binding. One hypothesis is that both exemplars being retrieved (PIP) provides the best measure of binding. We suggest that single items recalled under the paired condition may also reflect binding (Zimmer, Mecklinger, & Lindenberger, 2006) of the single item with the category cue. Compared to FNAME and STMB with cross-modal memory units (face vs. name, color vs. shape), the MBT involves units that are all semantic (two words for each category cue). Because facial recognition, color and shape processing may involve special regions of the brain (for example, fusiform face area for facial recognition; Weibert & Andrews, 2015), tests involving cross-modal memory units may be inherently more complex and less specific than the MBT that involves only semantic information.

The MBT may serve as a first-stage screening test to select individuals eligible for secondary prevention trials as well as biomarker and neuroimaging studies. Compared to neuroimaging biomarkers, the MBT has the advantages of being brief and inexpensive for screening. The cut-score choice may vary based on the particular public health context. When the goal is to select individuals for low-risk, low-cost secondary intervention trials, we may want to optimize the sensitivity; when the goal is to select individuals for more invasive treatments, we may want to optimize the specificity.

This study has some limitations. First, the sample sizes of the aMCI and dementia diagnostic groups were relatively small. In contrast to memory clinic samples that may include more aMCI and dementia patients, the smaller sample size of the dementia group in our study reflects the modest number of dementia cases in this population-based systematically recruited sample of older adults. The prevalent aMCI and dementia cases were the individuals who had not yet sought care for memory decline. Second, the aMCI and dementia cases were significantly older than the cognitively normal controls in this study. Because the current study was based on a population-based sample and age is a very important risk factor for dementia, the relatively older age observed in the aMCI and dementia groups was expected. Though the aMCI participants were older than healthy elderly controls in this study, their poor memory performance was far beyond the memory decline associated with normal aging based on the following results. For example, the FCSRT-FR scores were on average 13.4 points lower in the aMCI participants (19.4) than in normal controls (32.8) in this study. We had estimated an age-related annual decline of less than 0.30 point in FCSRT-FR scores using the longitudinal data from 1827 non-demented participants from the EAS (Mowrey et al. 2015). Given the age difference was on average less than 3 years between controls and aMCI in this study and annual change in FCSRT-FR was less than 0.30, which was a total of less than 1 point difference for 3 years and much smaller than the 13.4 points difference observed

between aMCI and control groups in this study. Third, we cannot compare the performance of MBT to that of FCSRT-FR and LM-I since FCSRT-FR and LM-I were used in the diagnostic criteria of this sample. However, the MBT binding measure TIP had better discriminative validity than the single list measure CR-L1, indicating better operating characteristics of the MBT in discriminating aMCI and dementia than memory tests of a single list format such as FCSRT. Fourth, we presented results on the immediate cued recall measures from the MBT. The MBT may have three additional steps which are immediate free recall, 30-min delayed free recall and delayed cued recall (Gramunt et al., 2015; Papp et al., 2015; Rentz et al., 2010). In this study, we did not collect data on these measures. Nonetheless, immediate cued recall is the least challenging recall condition of the test and these measures have achieved good reliability and good discriminative validity, indicating that the MBT immediate recall measures are both challenging and effective. The Harvard Aging Brain study group demonstrated that MBT delayed free recall of second list was associated with amyloid burden in precuneus in clinical normal elderly (Rentz et al., 2010). They have also shown that MBT delayed free recall but not cued recall was reduced in the prodromal biomarker stage with amyloid burden without neurodegeneration (Aβ+/ND−) while both delayed free and cued recall were reduced in the stage with amyloid burden and neurodegeneration (Aβ+/ND+) (Papp et al., 2015).

In conclusion this study has demonstrated that the MBT has good reliability and good discriminative validity to distinguish aMCI and dementia. These conclusions have good generalizability since our study is based on a population-based cohort study where covariates are systematically measured. Semantic memory binding, measured by the MBT, may emerge as a promising behavioral marker for AD early detection along with other associative binding tests such as STMB and FNAME (Amariglio et al., 2012; Parra et al., 2009; Rentz et al., 2011, 2013).

Supplementary Material

Supplementary material is available at *Archives of Clinical Neuropsychology* online.

Funding

Conflict of Interest

None declared.

Acknowledgements

References

Amariglio, R. E., Frishe, K., Olson, L. E., Wadsworth, L. P., Lorius, N., Sperling, R. A., et al. (2012). Validation of the face name associative memory exam in cognitively normal older individuals. *Journal of Clinical and Experimental Neuropsychology*, *34* (6), 580–587. doi:10.1080/13803395.2012.666230.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders, DSM-IV*. Washington, DC: American Psychiatric Association.

Artero, S., Petersen, R., Touchon, J., & Ritchie, K. (2006). Revised criteria for mild cognitive impairment: Validation within a longitudinal population study. *Dementia and Geriatric Cognitive Disorders*, *22* (5–6), 465–470. doi:10.1159/000096287.

Benton, A. L., & Hamsher, K. (1989). *Multilingual aphasia examination*. Iowa City: AJA Assoc.

Blessed, G., Tomlinson, B. E., & Roth, M. (1968). The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *The British Journal of Psychiatry: The Journal of Mental Science*, *114* (*512*), 797–811.

Buschke, H. (1984). Cued recall in amnesia. *Journal of Clinical Neuropsychology*, *6* (*4*), 433–440.

Buschke, H. (2014). Rationale of the Memory Binding Test. In Nilsson, L., & Ohta, N.(Eds.)*Dementia and memory* (pp.55–71). Hove, East Sussex: Psychology Press.

Chui, H. C., Victoroff, J. I., Margolin, D., Jagust, W., Shankle, R., & Katzman, R. (1992). Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's disease diagnostic and treatment centers. *Neurology*, *42* (*3 Pt 1*), 473–480.

Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., Dekosky, S. T., Barberger-Gateau, P., et al. (2010). Revising the definition of Alzheimer's disease: A new lexicon. *The Lancet Neurology*, *9* (*11*), 1118–1127. doi:10.1016/s1474-4422(10)70223-4.

Gramunt, N., Buschke, H., Sanchez-Benavides, G., Lipton, R. B., Pena-Casanova, J., Dieguez-Vide, F., et al. (2015). Reference data of the Spanish Memory Binding Test in a midlife population from the ALFA STUDY (Alzheimer's and family). *Journal of Alzheimer's Disease: JAD*, *48* (*3*), 613–625. doi:10.3233/jad-150237.

Gramunt, N., Sanchez-Benavides, G., Buschke, H., Dieguez-Vide, F., Pena-Casanova, J., Masramon, X., et al. (2016). The Memory Binding Test: Development of two alternate forms into Spanish and Catalan. *Journal of Alzheimer's Disease: JAD*. doi:10.3233/jad-151175.

Gramunt, N., Sanchez-Benavides, G., Buschke, H., Lipton, R. B., Masramon, X., Gispert, J. D., et al. (2016). Psychometric properties of the Memory Binding Test: Test-retest reliability and convergent validity. *Journal of Alzheimer's disease : JAD*, *50* (*4*), 999–1010. doi:10.3233/jad-150776.

Grober, E., & Buschke, H. (1987). Genuine memory deficits in dementia. *Developmental Neuropsychology*, *3* (*1*), 13–36.

Grober, E., & Kawas, C. (1997). Learning and retention in preclinical and early Alzheimer's disease. *Psychology and Aging*, *12* (*1*), 183–188.

Grober, E., Lipton, R. B., Hall, C., & Crystal, H. (2000). Memory impairment on free and cued selective reminding predicts dementia. *Neurology*, *54* (*4*), 827–832.

Jack, , C. R.,Jr, Knopman, D. S., Weigand, S. D., Wiste, H. J., Vemuri, P., Lowe, V., et al. (2012). An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease. *Annals of Neurology*, *71* (*6*), 765–775. doi:10.1002/ana.22628.

Jorm, A. F., & Jacomb, P. A. (1989). The informant questionnaire on cognitive decline in the elderly (IQCODE): Socio-demographic correlates, reliability, validity and some norms. *Psychological Medicine*, *19* (*4*), 1015–1022.

Julian, L. J., Gregorich, S. E., Earnest, G., Eisner, M. D., Chen, H., Blanc, P. D., et al. (2009). Screening for depression in chronic obstructive pulmonary disease. *COPD*, *6* (*6*), 452–458. doi:10.3109/15412550903341463.

Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). *The Boston naming test*. Philadelphia: Lea and Febiger.

Katz, M. J., Lipton, R. B., Hall, C. B., Zimmerman, M. E., Sanders, A. E., Verghese, J., et al. (2012). Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and Alzheimer dementia in blacks and whites: A report from the Einstein Aging Study. *Alzheimer Disease and Associated Disorders*, *26* (*4*), 335–343. doi:10.1097/WAD.0b013e31823dbcfc.

Katzman, R., Brown, T., Thal, L. J., Fuld, P. A., Aronson, M., Butters, N., et al. (1988). Comparison of rate of annual change of mental status score in four independent studies of patients with Alzheimer's disease. *Annals of Neurology*, *24* (*3*), 384–389. doi:10.1002/ana.410240306.

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *The Gerontologist*, *9* (*3*), 179–186.

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

Loewenstein, D. A., Acevedo, A., Schram, L., Ownby, R., White, G., Mogosky, B., et al. (2003). Semantic interference in mild Alzheimer disease: Preliminary findings. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, *11* (*2*), 252–255.

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*), 939–944.

McKhann, G. M., Albert, M. S., Grossman, M., Miller, B., Dickson, D., & Trojanowski, J. Q. (2001). Clinical and pathological diagnosis of frontotemporal dementia: Report of the Work Group on Frontotemporal Dementia and Pick's Disease. *Archives of Neurology*, *58* (*11*), 1803–1809.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., et al. (1989). The consortium to establish a registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39* (*9*), 1159–1165.

Mowrey, W., Grober, E., Zimmerman, M.E., Katz, M.J., Hall, C.B., Sliwinski, M.J., et al. (2015). Estimating episodic memory within-subject decline among non-demented older adults: Results from the Einstein Aging Study (EAS). *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *11* (*7*), 714.

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., et al. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51* (*6*), 1546–1554.

Papp, K. V., Amariglio, R. E., Mormino, E. C., Hedden, T., Dekhytar, M., Johnson, K. A., et al. (2015). Free and cued memory in relation to biomarker-defined abnormalities in clinically normal older adults and those at risk for Alzheimer's disease. *Neuropsychologia*, *73*, 169–175. doi:10.1016/j.neuropsychologia.2015.04.034.

Parra, M. A., Abrahams, S., Fabi, K., Logie, R., Luzzi, S., & Della Sala, S. (2009). Short-term memory binding deficits in Alzheimer's disease. *Brain: a Journal of Neurology*, *132* (*Pt 4*), 1057–1066. doi:10.1093/brain/awp036.

Parra, M. A., Abrahams, S., Logie, R. H., Mendez, L. G., Lopera, F., & Della Sala, S. (2010). Visual short-term memory binding deficits in familial Alzheimer's disease. *Brain: a Journal of Neurology*, *133* (*9*), 2702–2713. doi:10.1093/brain/awq148.

Rabin, L. A., Smart, C. M., Crane, P. K., Amariglio, R. E., Berman, L. M., Boada, M., et al. (2015). Subjective cognitive decline in older adults: An overview of self-report measures used across 19 international research studies. Journal of Alzheimer's Disease: JAD, *48* (*Suppl 1*), S63–86. doi: 10.3233/jad-150154.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276. doi:10.2466/PMS.8.7.271-276.

Rentz, D. M., Amariglio, R. E., Becker, J. A., Frey, M., Olson, L. E., Frishe, K., et al. (2011). Face-name associative memory performance is related to amyloid burden in normal elderly. *Neuropsychologia*, *49* (*9*), 2776–2783. doi:10.1016/j.neuropsychologia.2011.06.006.

Rentz, D. M., Locascio, J. J., Becker, J. A., Moran, E. K., Eng, E., Buckner, R. L., et al. (2010). Cognition, reserve, and amyloid deposition in normal aging. *Annals of Neurology*, *67* (*3*), 353–364. doi:10.1002/ana.21904.

Rentz, D. M., Parra Rodriguez, M. A., Amariglio, R., Stern, Y., Sperling, R., & Ferris, S. (2013). Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: A selective review. *Alzheimer's Research & Therapy*, *5* (*6*), 58. doi:10.1186/alzrt222.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77. doi:10.1186/1471-2105-12-77.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, *7* (*3*), 280–292. doi:10.1016/j.jalz.2011.03.003.

Steinberg, B. A., Bieliauskas, L. A., Smith, G. E., & Ivnik, R. J. (2005). Mayo's older Americans normative studies: Age- and IQ-adjusted norms for the Wechsler memory scale—Revised. *The Clinical Neuropsychologist*, *19* (*3–4*), 378–463. doi:10.1080/13854040590945201.

Villeneuve, S., & Belleville, S. (2012). The nature of memory failure in mild cognitive impairment: examining association with neurobiological markers and effect of progression. *Neurobiology of Aging*, *33* (*9*), 1967–1978. doi:10.1016/j.neurobiolaging.2011.10.004.

Wechsler, D. (1987). *Wechsler memory scale—Revised manual*. San Antonio: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). New York: The Psychological Corporation.

Weibert, K., & Andrews, T. J. (2015). Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces. *Neuropsychologia*, *75*, 588–596. doi:10.1016/j.neuropsychologia.2015.07.015.

Wilkinson, G. S. (1993). *The wide range achievement test: Manual* (3rd ed.). Wilmington, DE: Wide Range.

Zimmer, H. D., Mecklinger, A., & Lindenberger, U. (2006). *Handbook of binding and memory, perspective from cognitive neuroscience*. New York: Oxford University Press.