Published in final edited form as:

Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. 2009 May ; 16(3): 357–376. doi:
10.1080/13825580902825220.

Differential Memory Test Sensitivity for Diagnosing Amnestic Mild Cognitive Impairment and Predicting Conversion to Alzheimer's Disease

Laura A. Rabin^{1,2}, Nadia Paré^{2,3}, Andrew J. Saykin^{2,4}, Michael J. Brown¹, Heather A. Wishart², Laura A. Flashman², and Robert B. Santulli²

¹Department of Psychology, Brooklyn College and the Graduate Center of the City University of New York, Brooklyn, NY, 11210

²Brain Imaging Laboratory, Department of Psychiatry, Dartmouth Medical School, Lebanon, NH, 03756

³Department of Neuropsychology, Nebraska Medical Center, Omaha, NE, 68198

⁴Center for Neuroimaging, Department of Radiology, Indiana University School of Medicine, Indianapolis, IN, 46202

Abstract

Episodic memory is the first and most severely affected cognitive domain in Alzheimer's disease (AD), and it is also the key early marker in prodromal stages including amnestic mild cognitive impairment (MCI). The relative ability of memory tests to discriminate between MCI and normal aging has not been well characterized. We compared the classification value of widely used verbal memory tests in distinguishing healthy older adults (n = 51) from those with MCI (n = 38).

Univariate logistic regression indicated that the total learning score from the California Verbal Learning Test-II (CVLT-II) ranked highest in terms of distinguishing MCI from normal aging (sensitivity = 90.2; specificity = 84.2). Inclusion of the delayed recall condition of a story memory task (i.e., WMS-III Logical Memory, Story A) enhanced the overall accuracy of classification (sensitivity = 92.2; specificity = 94.7). Combining Logical Memory recognition and CVLT-II long delay best predicted progression from MCI to AD over a 4-year period (accurate classification = 87.5%). Learning across multiple trials may provide the most sensitive index for initial diagnosis of MCI, but inclusion of additional variables may enhance overall accuracy and may represent the optimal strategy for identifying individuals most likely to progress to dementia.

INTRODUCTION

Episodic memory, defined as recollection of specific past events or information, is the first and most severely affected cognitive domain in Alzheimer's disease (AD) (Bäckman, Jones, Berger, Laukka, & Small, 2005). Episodic memory deficits are a key indicator of prodromal dementia stages, specifically for amnestic mild cognitive impairment (MCI) (de Jager & Budge, 2005; Ganguli, Dodge, Shen, & DeKosky, 2004; Kavé & Heinik, 2004; Marruff et al., 2004), and they are thought to represent the effect of early neuropathological changes to the hippocampal and entorhinal cortices (Becker et al., 2006; Petrella et al., 2006; Wicklund, Johnson, Rademaker, Weitner, & Weintraub, 2006). MCI is considered a transitional period

between normal aging and dementia (Morris, 2006; Petersen et al., 2001). MCI patients experience memory loss that differs significantly from normal age-related change, but these individuals do not meet criteria for dementia due to their generally preserved overall cognitive function and activities of daily living (ADLs). Several studies have reported episodic memory impairment in preclinical AD anywhere from 3 to 8 years prior to formal diagnosis of AD (Amieva et al., 2005; Bäckman, Small, & Fratiglioni, 2001; Saxton et al., 2004).

The usefulness of neuropsychological measures for detecting early cognitive change is well established (Arnaiz & Almkvist, 2003), with some studies reporting accuracy rates of 85 – 90% for correct identification of MCI patients who will progress to AD (Collie & Marruff, 2000; Salmon et al., 2002; Small, Fratiglioni, Viitanen, Winblad, & Bäckman, 2000). With such high accuracy, however, comes variable levels of sensitivity and specificity for individual neuropsychological instruments as predictors of diagnostic conversion. Modrego (2006) reported that sensitivity and specificity varied from 64 to 89% and 76 to 97% respectively, depending on the measure of cognitive status utilized. In addition to this variability, memory measures have limitations when used to distinguish normal aging from MCI (Collie & Marruff, 2000). Widely used screening tests such as the Mini Mental Status Examination (MMSE) and Mattis Dementia Rating Scale-2 (DRS-2) have a ceiling effect in MCI, making it difficult to distinguish normal aging from early cognitive decline when using these measures (de Jager, Hogervorst, Combrinck, & Budge, 2003). Practice effects are another potential limitation, masking the detection of true serial change over time in MCI (McCaffrey, Duff, & Westervelt, 2000). Another factor that confounds the interpretation of neuropsychological test performance is the standardization process. Normative samples can be distorted by standardization procedures that fail to consider risk factors for vascular disease and dementia, serving to lower mean performance levels and increase variability-- resulting in an unreliable estimate of the “normal” range of functioning on certain tests (Brooks, Iverson, Holdnack, & Feldman, 2008; de Jager et al., 2003; Gurol et al., 2006; Holtzer et al., 2008; Ritchie, Frerichs, & Tuokko, 2007). These confounding variables (i.e., ceiling effects, practice effects, standardization limitations) make it difficult to identify appropriate memory measures that are both sensitive and specific for detection of MCI.

Arnaiz and Almkvist (2003) reviewed the literature and reported that verbal episodic memory was generally considered the best predictor of cognitive decline in MCI, though instruments utilized across studies were not homogenous and likely differed in terms of sensitivity. Moreover, differential results appeared to arise from variation in task difficulty in memory tests rather than from variations in the specific processes that the tasks intended to measure. In another study comparing different forms of episodic memory measures, Ivanoiu and colleagues (2005) found high specificity (100%) but extremely low sensitivity (12%) for the MMSE in the diagnosis of MCI. Even though patients attained significantly lower scores on both the MMSE and the DRS-Memory tests relative to healthy controls, both measures failed to differentiate declining versus stable patients and therefore were not helpful in distinguishing those at risk for conversion to AD. Others have found similar results with incipient AD patients (Karrasch, Sinerva, Gronholm, Rinne, & Laine, 2005; Tierney, Szalai, Dunn, Geslani, & McDowell, 2000), highlighting the importance of using instruments that sufficiently tax memory resources.

List learning tasks, which involve learning across multiple trials, are perhaps the most challenging and sensitive measures of episodic memory when used to assess early cognitive changes in MCI and AD. De Jager and colleagues (2003) found that compared to story recall and a visual memory test, only the total learning score of the Hopkins Verbal Learning Test (HVLT) (Brandt, 1991) discriminated MCI patients and healthy controls. Kavé and Heinik

(2004) found that study participants met criteria for MCI on delayed recall when assessed using a list learning test (Rey-Auditory Verbal Learning Test, RAVLT; Schmidt, 1996), though they demonstrated normal performance on a visual memory test (Rey-Osterrieth Complex Figure Test; Spreen & Strauss, 1988) and story recall (Logical Memory I and II from the Wechsler Memory Scale-Third Edition, WMS-III; Wechsler, 1997). The authors concluded that in order to make a diagnosis of MCI, a marked decline in at least one memory test, most likely a list learning task, should be obtained. They also argued that requiring impairment on all memory measures would probably lead to a delay in diagnosis until the criteria for dementia had already been met, in turn skipping the MCI stage altogether.

In addition to providing strong diagnostic sensitivity for MCI, list learning tests can accurately predict diagnostic conversion to AD (Griffith et al., 2006; Marruff et al., 2004) and help characterize the severity of the memory impairment in early stages of dementia (Fox, Olin, Erbllich, Ghosh Ippen, & Schneider, 1998). Karrasch and colleagues (2005) reported modest sensitivity (33%) and high specificity (93%) for the 10-item word list test from the Consortium to Establish a Registry in Alzheimer's Disease (CERAD; Morris et al., 1989) battery in MCI. The authors also noted that utilizing a test with a greater number of words should lead to increased sensitivity. Tierney, Yao, Kiss, and McDowell (2005) found that within a comprehensive dementia work-up that included a neuropsychological assessment, the short delayed recall of the 15-item RAVLT was the only measure to emerge from a 10-year regression analysis as a predictor of AD; the RAVLT also emerged as the most significant test in the 5-year prediction analysis. Esteves-Gonzalez, Kulisevsky, Boltes, Otermin, and Garcia-Sanches (2003) found that the total learning and delayed recall scores of the RAVLT helped identify individuals with subjective memory complaints who would progress to AD over at least a 1-year interval, and it also differentiated between MCI and normal aging. It should be noted that other memory tests have shown predictive validity for clinical progression from MCI to AD, including delayed recall of story memory and verbal paired associates (Guarch, Marcos, Salamero, Gastó, & Blesa, 2008) and visuospatial paired associate learning (Ahmed, Mitchell, Arnold, Nestor, & Hodges, 2008), but to our knowledge these measures have not been directly compared to list learning tasks.

The current study compared the ability of widely used clinical memory tests to classify individuals as either MCI or healthy controls (HC). We also sought to determine whether a single memory variable or combination of variables would provide the greatest diagnostic sensitivity and specificity for MCI. Finally, in our cohort of MCI participants, we investigated which memory test(s) accurately predicted conversion from MCI to probable AD over a follow-up period.

METHODS

Participants were recruited into a longitudinal memory and aging study through flyers, public lectures, newspaper advertisements, and referrals from various Medical Center clinics. Screening for eligibility involved review of recent medical records and standardized phone interview, which included basic demographic information, medical and psychiatric history, and objective and subjective memory screens (Rabin et al., 2007). Inclusion criteria were age 60 years or greater, right-handedness, and fluency in English. To be consistent with the population characteristics from which this sample was recruited, participants were required to have at least 12 years of formal education or a GED. Exclusion criteria included medical, psychiatric, or neurological conditions (other than MCI) that could significantly affect brain structure or cognitive status; head injury with loss of consciousness greater than five minutes; history of cancer with chemotherapy and/or radiation therapy; use of psychotropic medication; and current or past substance dependence. Also excluded were

individuals with non-amnesic forms of MCI (Petersen, 2004; Winblad et al., 2004) as determined on the basis of a multidisciplinary, clinical consensus meeting that included neuropsychological test findings and other relevant data. Despite active attempts to recruit minority participants, the sample was ultimately composed entirely of Caucasian individuals, consistent with the predominant demographic composition of the surrounding northern New England region. All participants provided written informed consent according to procedures approved by the Institutional Committee for the Protection of Human Subjects.

Approximately eight weeks after telephone screening, eligible participants underwent a comprehensive neuropsychological evaluation, including measures of memory, attention, language, executive functions, visuospatial ability, general intellectual ability, and psychomotor speed, as well as standard dementia screens. Tests were administered by postdoctoral fellows and highly trained research assistants who were blind to diagnostic status at baseline assessment. All tests were administered during a single session at our medical center, with a fixed order of administration designed to minimize the introduction of any new verbal stimuli during delay periods for verbal memory tests. Tests included: Action Naming Test (Piatt, Fields, Paolo, & Troster, 1999; 2004), Barona Index (Barona, Reynolds, & Chastain, 1984), Boston Naming Test (Goodglass, Kaplan, & Barresi, 2001), California Verbal Learning Test, Second Edition (CVLT II; Delis, Kramer, Kaplan, & Ober, 2000), Mattis Dementia Rating Scale-2 (DRS-II; Jurica, Leitten, & Mattis, 2001), Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), Test of Practical Judgment (Rabin et al., 2007), Trail Making Test (Delis, Kaplan, & Kramer, 2001; Reitan & Wolfson, 1985), Wechsler Adult Intelligence Scale-Third Edition (WAIS-III Digit Symbol, Digit Span, Block Design, Vocabulary, and Information subtests; Wechsler, 1997), Wechsler Memory Scale-Third Edition (WMS-III Logical Memory [LM] I and II and Visual Reproduction [VR] I and II subtests and LM and VR Recognition; Wechsler, 1997), and Wisconsin Card Sorting Test (short form; Heaton, Chelune, Talley, Kay, & Curtis, 1993). Estimates of baseline intellectual functioning included WAIS-III Vocabulary and the American National Adult Reading Test (Grober & Sliwinski, 1991).

Participants and their informants also completed a series of cognitive questionnaires. A Cognitive Complaint Index (CCI; Saykin et al., 2006) ranging from 0 to 100 was calculated based on the total number of items that could be endorsed across all subjective measures including: Memory Self-Rating Questionnaire (Squire, Wetzel, & Slater, 1979), self- and informant- versions of the Neurobehavioral Function and Activities of Daily Living Rating Scale (Saykin, 1992; Saykin et al., 1991), self- and informant-versions of the Informant Questionnaire on Cognitive Decline in the Elderly (Jorm & Jacomb, 1989), four cognitive items from the Geriatric Depression Scale (GDS; Yesavage et al., 1982), 10 cognitive items from a telephone-based screening for MCI (Rabin et al., 2007) and 23 items from the Memory Assessment Questionnaire (Santulli et al., 2005), adapted in part from the Functional Activities Questionnaire (Pfeffer et al., 1982). A Board-certified geropsychiatrist ruled out significant impairment in activities of daily living (ADLs), depression, and other Axis I psychiatric conditions based on a semi-structured interview that included the Hamilton Rating Scale for Depression (Hamilton, 1960). A neurological examination and the original and revised Hachinski Ischemia Scales (Hachinski et al., 1975; Small, 1985) were completed to rule out incidental pathology.

Criteria for MCI were based on the Petersen recommendations (Peterson, 2004; Petersen et al. 2001), operationalized for the purpose of this study: 1 (1) Presence of significant cognitive complaints, preferably corroborated by an informant. Specifically, we required endorsement of > 20% of possible complaints on the CCI. (2) Presence of objective memory impairment. Specifically, consistent with the methodology utilized by Troyer et al. (2008),

we required that the age-corrected scaled score (age SS) on one or more memory tests be considerably lower (i.e., ≥ 1.5 standard deviations lower) than was the age SS for Verbal IQ (as measured by WAIS-III Vocabulary, mean = 10, SD = 3). The standardized memory tests selected for classification purposes were WMS-III Visual Reproduction Immediate Recall, Delayed Recall, and Recognition conditions and the DRS-2 memory subtest. The memory tests were selected to avoid criterion contamination of the verbal memory measures of particular interest for the present report and were not used in the logistic regression analyses. (3) Unimpaired activities of daily living as determined by a multidisciplinary clinical consensus team, incorporating data from the geropsychiatric evaluation, telephone screen, and Neurobehavioral Function and Activities of Daily Living Rating Scale. (4) Preserved general cognitive functioning as measured by cutoff scores of > 24 of 30 or ≥ 128 of 144 on the MMSE and DRS-2 respectively. (5) No clinical depression or other psychiatric disorder as determined by a multidisciplinary clinical consensus team, and incorporating information from the geropsychiatric evaluation, Geriatric Depression Scale, and Hamilton Rating Scale for Depression.

Healthy control participants had: (1) No significant cognitive complaints, as indicated by endorsement of $\leq 20\%$ of possible complaints on the CCI. (2) Intact functioning on memory testing. Specifically, we required that the age-corrected scaled score (age SS) for all of the following memory tests not fall considerably (i.e., ≥ 1.5 standard deviations) below the age SS for Verbal IQ (as measured by WAIS-III Vocabulary, mean = 10, SD = 3): WMS-III Visual Reproduction Immediate Recall, Delayed Recall, and Recognition and DRS-2 memory. (3) Normal activities of daily living measured as described for the MCI group. (4) Preserved general cognitive functioning also measured as described for the MCI group. (5) No clinical depression or other psychiatric disorder as described for the MCI group. As noted above, participants who demonstrated impairment in cognitive domains other than memory were not eligible for study inclusion. We also analyzed the baseline scores of individuals who later went on to develop AD. These participants met criteria for probable mild AD, as defined by the NINCDS-ADRDA criteria (McKhann et al., 1984), over the course of longitudinal follow-up.

Memory tests and Statistical Analyses

Tests selected for the regression equations encompassed a range of memory processes, task procedures, types of stimuli, and difficulty levels. For example, we included the three word memory from the MMSE and two list learning tasks: the 16-item California Verbal Learning Test-Second Edition (CVLT-II) with five learning trials and short- and long-delay recall conditions as well as a 10-item list learning test developed in our lab for purposes of telephone screening with three learning trials and a delayed recall condition (Memory and Aging Telephone Screen, MATS; Rabin et al., 2007). As a measure of story recall, we used the WMS-III Logical Memory (LM) subtest with immediate, delayed recall, and recognition conditions. We also included delayed recall of LM Story A, as an earlier version of this measure is currently used as part of the diagnostic criteria for amnesic MCI in the Alzheimer's Disease Neuroimaging Initiative (Mueller et al., 2005) and the Alzheimer's Disease Cooperative Study (ADCS) clinical trials (Thal, 2005). Finally, we used WAIS-III Information (INFO), considered a test of remote memory or longer-term recall in older adult populations (Lezak, Howieson, & Loring, 2004), which has been shown to be predictive of AD within 5 years of diagnosis (Tierney et al., 2005). Notably, we were unable to include

¹Our procedure for diagnosing MCI in the current study differs from our standard approach reported elsewhere (e.g., Saykin et al., 2004, 2006; Rabin et al., 2007), in which classification decisions are made by a multidisciplinary clinical consensus panel, applying criteria developed by Petersen and colleagues and incorporating all neuropsychological test data and other available information. In order to avoid criterion contamination in the current study, we rigorously applied these criteria and used only the measures described herein to assign participants to groups.

WMS-III Visual Reproduction or DRS-2 memory subtests because these tasks were used to classify participants.

Logistic regression was employed to determine the maximal sensitivity and specificity of each test and to permit a ranking of tests that were considered to be most suitable for group differentiation. Multivariate logistic regression was used to calculate the optimal test combination for group discrimination. Age, education, and sex were included as covariates in all statistical analyses.

RESULTS

Participants included 89 older adults who met the eligibility criteria specified above and were classified as either MCI ($n = 38$) or HC ($n = 51$). The average number of memory measures (out of five possible) on which MCI participants fell below the mean was 2.48, $SD = .99$. There were no significant group differences in demographics (see Table 1). As mentioned above, age, education, and sex were all entered as covariates in the analysis and did not significantly affect the pattern of results.

We used univariate logistic regression to calculate the optimal sensitivity and specificity for each test (see Table 2). The CVLT-II learning score was ranked highest (sensitivity = 90.2; specificity = 84.2) in distinguishing MCI from normal aging. LM recognition (LMREC) and CVLTII long delay scores ranked second and third, respectively, in terms of discriminability. LM immediate recall (LM I) and INFO emerged lower on the list, and attained relatively low specificity. A summary of the regression analysis for each test is provided in Table 3. Next, we used multivariate logistic regression to calculate the optimal combination of tests for group discrimination. Combining CVLT-II learning and delayed recall of LM Story A (LM II-A) ($-2 \text{ Log Likelihood} = 27.68$; Nagelkerke $R^2 = .88$) yielded the highest overall classification (sensitivity = 92.2; specificity = 94.7). Combining CVLT-II learning and MATS delay recall ($-2 \text{ Log Likelihood} = 30.26$; Nagelkerke $R^2 = .82$) yielded the lowest overall classification (sensitivity = 88.4; specificity = 88.2).

We also used univariate logistic regression to determine which test(s) predicted progression from MCI to probable AD. As mentioned previously, these participants met criteria for probable mild AD, as defined by the NINCDS-ADRDA criteria (McKhann et al., 1984).² Only MCI participants with at least one wave of follow-up data ($n = 32$) were included in this analysis. The six MCI participants who were lost to follow-up were statistically comparable to the remaining 32 in terms of baseline age, years of education, sex, and performance across memory tests: CVLT-II total, CVLT-II short delay, CVLT-II long delay, LMI, LMII, LMREC, MATS Total, MATS delay, MMSE W-Missed, and INFO ($p > .05$ for all analyses). The average duration of follow-up was 2.97 yrs, $SD = 1.16$ (range = 1 – 4 years). Nine participants from the MCI group developed probable AD (27.3%), resulting in an MCI comparison group of 23 participants. The mean time of diagnostic conversion for the nine participants was 1.89 yrs, $SD = 1.05$ (range = 1 – 4 years). As expected based on prior research (DeCarli et al., 2004; Howieson et al., 2008; Karrasch et al., 2005; Petersen & Negash, 2007), MCI participants who underwent diagnostic conversion showed significant performance reductions on memory tests at baseline as compared to their MCI counterparts who did not convert to AD during the observation period (see Table 4).

Results indicated that LMREC (sensitivity = 95.7; specificity = 62.5) and CVLT-II long delay (sensitivity = 100; specificity = 44.4) were the highest ranked tests in terms of

²Examiners could not always be blinded at follow-up but the clinical ratings were adjudicated by a multidisciplinary consensus conference comprised of clinicians and researchers, some of whom had had contact with the participant and some of whom had not.

distinguishing converters from non-converters (see Table 5). LM II-A ranked third in terms of discriminability. Notably, specificity was low across memory tests in this analysis (range 0% to 62.5%). A summary of the regression analysis for each test is provided in Table 6. Next, we used multivariate logistic regression to determine which test combination would yield the highest group discrimination. Combining LMREC and CVLT-II long delay (-2 Log Likelihood = 26.10; Nagelkerke $R^2 = .45$) resulted in increased specificity (66.7); however, there was no meaningful change in sensitivity (97.5) or accurate classification (87.5). Combining LMREC with LM I (-2 Log Likelihood = 25.64; Nagelkerke $R^2 = .46$) yielded the lowest overall group classification (sensitivity = 62.6; specificity = 55.6; accurate classification = 75.0).

DISCUSSION

This study compared various memory tests in terms of their ability to classify older adults as MCI or healthy controls. Results indicated that the CVLT-II total learning score across five trials provided the most accurate classification of participants, followed closely by recognition and delayed recall trials of LM and the CVLT-II, respectively. This finding is consistent with prior reports regarding the usefulness of list learning tasks for detecting early AD (Estevez-Gonzalez, 2003; Tierney et al., 2005) and with Petersen's observation that a sensitive index of MCI is the inability to acquire information beyond the short term memory span, and that this deficit becomes apparent over several learning trials (Petersen, Smith, Ivnik, Kokmen, & Tangalos, 1994). Not surprisingly, total learning score from a shorter, 10-item list learning task (Rabin et al., 2007) did not provide as strong diagnostic sensitivity or specificity as the 16-item CVLT-II. This is consistent with the idea that sensitivity of word list learning tests may be enhanced by increasing both the number of items to be remembered (> 10 items) and number of learning trials (> 3 trials) (Karrasch et al., 2005), while exerting caution to ensure that task demands are not too overwhelming for the population of interest. It is also important to note that number of words missed on the MMSE, immediate recall of a story memory test (i.e., WMS-III LM I), and remote memory (as measured by WAIS-III Information) were among the lowest ranked tests in terms of discriminability. These findings should be borne in mind when designing neuropsychological test batteries for clinical or research purposes.

We also sought to determine whether a single memory test or combination of variables would provide the greatest diagnostic sensitivity and specificity for MCI. Inclusion of a delayed recall condition of a story memory task (i.e., WMS-III LM Story A) enhanced the overall accuracy of classification (from 87.6% to 93.3%). This finding is interesting in light of controversy regarding the sensitivity of story recall versus list learning tests in detecting early memory changes. For example, some have suggested that story recall should be paired with a more demanding, list learning task to ensure proper detection of MCI, which often presents as subtle memory change (Kavé & Heinik, 2004). Indeed, combining list learning with story memory could make up for the lack of sensitivity of story memory, generally considered a less taxing measure because it provides a logical context for the examinee. This may be especially obvious in assessments of highly educated or highly functioning individuals who can better compensate for their declining memory capacity. The strategy of combining tasks also might provide a more accurate assessment of MCI individuals who perform poorly on list learning tasks for reasons other than true memory deficits (e.g., they become overwhelmed by task demands). Another possibility would be to develop new, more challenging measures of story memory, which might be able to detect subtle memory decline on their own.

Another goal was to determine which memory test(s) best predicted conversion from MCI to AD. It should be noted that the sample size for these analyses was considerably smaller than

for the previous analyses, and results should be interpreted with caution given issues of statistical power and generalizability of findings. In our sample of 32 MCI participants with available follow-up data, 27% (n = 9) progressed to probable AD, consistent with published conversion rates (Busse, Hensel, Guhne, Angermeyer, & Riedel-Heller, 2006; Daly et al., 2000). Findings revealed that the LM recognition was the highest ranked test in terms of classification (87.1%) followed closely by the delayed recall condition of CVLT-II (84.4%) and LM-II A (81.3%). It is notable that four of the top five tests for predicting diagnostic conversion all were episodic memory tests of delayed recall or recognition as opposed to tests of learning or short-term recall. These findings are consistent with the idea that immediate recall trials may measure attention and working memory (extensively mediated by prefrontal cortex and related circuitry) to a greater degree than delayed-recall trials, which primarily tap retention and delayed-retrieval effects (and on which performance is highly dependent on entorhinal and hippocampal systems) (Shankle et al., 2005). These distinctions may account for the superiority of delayed recall or recognition tasks for predicting diagnostic conversion as compared to making the initial diagnosis. Additionally, previous research has shown that the delayed recall condition of a lengthy supraspan test (i.e., RAVLT) can predict conversion to AD (Esteves-Gonzalez et al., 2003). Our findings together with these observations make a compelling case for using a combination of story memory and list learning tasks to ensure accurate detection of MCI and the ability to monitor progressive decline.

Future research might expand the current findings to include memory paradigms not traditionally assessed during neuropsychological evaluations. For example, Troyer et al. (2008) reported on the clinical usefulness of calculating associative memory scores from standardized tests of object-location recall (i.e., BVMT-R) and symbol-symbol recall (i.e., Digit symbol incidental recall) as part of MCI evaluations. Specifically, their ROC curve analyses showed better classification of individuals to MCI or healthy control groups with measures of associative recall as opposed to item recall. Anderson et al. (2008) explored the roles of automatic and controlled memory mechanisms in older adults and found that individuals with MCI had intact familiarity but a marked difficulty recollecting prior items in their context. The authors suggested that this selective effect likely reflects disrupted hippocampal functioning and offered suggestions for rehabilitation focused on enhancing contextual recollection or training recollection directly.

Given issues of respondent burden and battery construction within our longitudinal study, we were limited to a relatively small group of memory tests. Further, to avoid circularity in defining and discriminating MCI, we used the WMS-III Visual Reproduction and DRS-2 Memory for diagnostic purposes and did not include these measures in the regression analyses. Ideally, in future research we would classify MCI and HC participants using memory tests with stronger diagnostic validity to minimize the possibility of misclassification. We would also strive to include additional measures to enable comparison of the relative value of visual versus verbal memory tests. Combining the CVLT-II and WMS-III LM with other types of memory tasks might lead to the optimal assessment method for detection of MCI and monitoring of clinical course. As well, it might be possible to enhance the diagnostic and predictive value of list learning tests by using techniques such as correspondence analysis, in which weighted scores are derived for each participant from their item responses on immediate- and delayed-recall trials. This method, which weights each word's relative importance by its position in the list and by the trial in which it was recalled, has proven superior to use of straightforward total learning or cutoff scores for early detection of AD (Shankle et al., 2005).

Generalization of our results is limited by the high level of education and estimated baseline level of functioning of the participants. Cognitive reserve theory (Spitznagel, Tremont,

Brown, & Gunstad, 2006; Stern, 2002) predicts that highly educated older adults will perform better for a longer period of time before the load of the medial temporal pathology impairs memory functioning. Thus, it will be important to replicate these findings in a more diverse group of elders who may demonstrate impairment on some of our less taxing memory measures (e.g., MMSE words missed). The current results are not meant to characterize non-amnestic forms of MCI, which may present with cognitive impairment in multiple domains or a single non-memory domain. Given the focus of the present report and power limitations, we were unable to consider non-memory variables in the analyses. Furthermore, although our sample was recruited on the basis of amnestic features of MCI and no participant demonstrated severe or disproportionate impairment on non-memory tests, it is widely accepted that some MCI participants will show subtle deficits on tests of other cognitive domains. For example, neuropsychological measures of executive function have been shown to be a prominent feature of prodromal AD and useful in some predictive models (Albert, Blacker, Moss, Tanzi, & McArdle, 2007; Albert, Moss, Tanzi, & Jones, 2001; Chen et al., 2000; Elias et al., 2000). Inclusion of executive function tests in our predictive models may have yielded valuable diagnostic information and improved the specificity of our measures for predicting dementia. Finally, although the combination of CVLT-II learning and LM Story A delayed recall offered the optimal group discrimination for our participant groups, all combinations of available tests yielded fairly high sensitivity (> 88%) and specificity (> 88%), implying that use of more than one memory measure is a key factor in enhancing classification accuracy (perhaps nearly as important as the specific choice of tests). It is likely that inclusion of multiple memory processes and the combined total number of memory items presented are important. Measurement reliability is related to total item count, which may be a key psychometric factor.

In summary, learning across multiple trials appears to be the most sensitive diagnostic index for MCI, though inclusion of additional variables can enhance overall accuracy. Further, though delayed recall tests best predicted progression from MCI to AD, the specificity of memory measures on their own was low and may be enhanced by using multiple tests covering various cognitive domains. Overall, it is clear that learning across multiple trials and memory for new supraspan information will continue to play an important role in research on preclinical dementia in terms of determining study eligibility, examining treatment effects, and monitoring the cognitive status of participants over time. To this end, it will be important to develop and utilize instruments with multiple, equivalent test forms to reduce practice effects or use empirical methods to determine whether observed change scores exceed chance expectations (Chelune, 2003). Future research also might investigate the usefulness of specific memory tests in relation to important biomarkers such as APOE epsilon 4 genotype status, hippocampal and entorhinal cortex volumes, CSF and plasma A β 42 concentrations, and PET amyloid imaging. These biomarkers are important for research but are expensive and ultimately better cognitive assessment protocols, if validated, could help reduce clinical diagnostic and monitoring costs for individuals known to be at risk for developing AD. Finally, list learning and story memory tests permit investigation into various aspects of memory functioning (e.g., encoding, retrieval, recognition, strategy use, the contribution of meaning to retention and recall). Future research should more closely examine which memory processes are tapped by specific tests and the nature and extent of decline in these processes during the insidious transition from healthy aging to MCI to AD.

Acknowledgments

Supported, in part, by grants to AJS from the National Institute on Aging (R01 AG19771) and the Alzheimer's Association (IIRG-99-1653 sponsored by the Hedco Foundation). The authors thank Margaret Nordstrom, Katherine Nutter-Upham, and Heather Pixley for their assistance with this study.

REFERENCES

- Albert M, Blacker D, Moss MB, Tanzi R, McArdle JJ. Longitudinal change in cognitive performance among individuals with mild cognitive impairment. *Neuropsychology*. 2007; 21:158–169. [PubMed: 17402816]
- Albert MS, Moss MB, Tanzi R, Jones K. Preclinical prediction of AD using neuropsychological tests. *Journal of the International Neuropsychological Society*. 2001; 7:631–639. [PubMed: 11459114]
- Amieva H, Jacqmin-Gadda H, Orgogozo J-M, Le Carret N, Helmer C, Letenneur L, Barberger-Gateau P, Fabigoule C, Dartigues J-F. The 9-year cognitive decline before dementia of the Alzheimer type: A prospective population-based study. *Brain*. 2005; 128:1093–1101. [PubMed: 15774508]
- Arnaiz E, Almkvist O. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. *Acta Neurologica Scandinavica*. 2003; 107:34–41.
- Arnold S, Mitchell J, Arnold R, Nestor PJ, Hodges JR. Predicting rapid clinical progression in amnesic mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*. 2008; 25:170–177. [PubMed: 18212499]
- Bäckman L, Jones S, Berger A-K, Laukka EJ, Small B. Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*. 2005; 19:520–531. [PubMed: 16060827]
- Bäckman L, Small BJ. Influences of cognitive support on episodic remembering: Tracing the process of loss from normal aging to Alzheimer's disease. *Psychology and Aging*. 1998; 13:267–276. [PubMed: 9640587]
- Bäckman L, Small BJ, Fratiglioni L. Stability of the preclinical episodic memory deficits in Alzheimer's disease. *Brain*. 2001; 124:96–102. [PubMed: 11133790]
- Barona A, Reynolds C, Chastain R. A demographically based index of pre-morbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology*. 1984; 52:885–887.
- Becker JT, Davis SW, Hayashi KM, Meltzer CC, Toga AW, Lopez OL, Thompson PM. Three-dimensional patterns of hippocampal atrophy in mild cognitive impairment. *Archives of Neurology*. 2006; 63:97–101. [PubMed: 16401741]
- Brandt J. The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *Clinical Neuropsychologist*. 1991; 5:125–142.
- Brooks BL, Iverson GL, Holdnack JA, Feldman HH. Potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society*. 2008; 14:463–478. [PubMed: 18419845]
- Busse A, Hensel A, Guhne U, Angermeyer MC, Riedel-Heller SG. Mild cognitive impairment: Long-term course of four clinical subtypes. *Neurology*. 2006; 67:2176–2185. [PubMed: 17190940]
- Chelune, GJ. Assessing reliable neuropsychological change. In: Franklin, RD., editor. *Prediction in forensic and neuropsychology: Sound statistical practices*. Lawrence Erlbaum Associates Publishers; Mahwah, NJ, US: 2003. p. 123-147.
- Chen P, Ratcliff G, Belle SH, Cauley JA, DeKosky ST, Ganguli M. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology*. 2000; 55:1847–1853. [PubMed: 11134384]
- Collie A, Marruff P. The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment. *Neuroscience and Biobehavioral Reviews*. 2000; 24:365–374. [PubMed: 10781696]
- Craft S, Newcomer J, Kanne S, Dagogo-Jack S, Cryer P, Sheline Y, et al. Memory improvement following induced hyperinsulinemia in Alzheimer's disease. *Neurobiology of Aging*. 1996; 17:123–130. [PubMed: 8786794]
- Daley E, Zaitchik D, Copeland M, Schmahmann J, Gunther J, Albert M. Predicting conversion to Alzheimer's disease using standardized clinical information. *Archives of Neurology*. 2000; 57:675–680. [PubMed: 10815133]
- DeCarli C, Mungas D, Harvey D, Reed B, Weiner M, Chui H, Jagust W. Memory impairment, but not cerebrovascular disease, predicts progression of MCI to dementia. *Neurology*. 2004; 63:220–227. [PubMed: 15277612]
- Delis, DC.; Kramer, JH.; Kaplan, E.; Ober, BA. *California Verbal Learning Test-Second Edition: Adult version manual*. Psychological Corporation; San Antonio, TX: 2000.

- Delis, D.; Kaplan, E.; Kramer, JH. Delis-Kaplan Executive Function System. Psychological Corporation; San Antonio, TX: 2001.
- de Jager CA, Budge MM. Stability and predictability of the classification of mild cognitive impairment as assessed by episodic memory test performance over time. *Neurocase*. 2005; 11:72–79. [PubMed: 15804927]
- de Jager CA, Hogervorst E, Combrinck M, Budge MM. Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine*. 2003; 33:1039–1050. [PubMed: 12946088]
- Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB. The preclinical phase of Alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Archives of Neurology*. 2000; 57:808–813. [PubMed: 10867777]
- Estevez-Gonzalez A, Kulisevsky J, Boltes A, Otermin P, Garcia-Sanchez C. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *International Journal of Geriatric Psychiatry*. 2003; 18:1021–1028. [PubMed: 14618554]
- Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198. [PubMed: 1202204]
- Fox LS, Olin JT, Erblieh J, Ghosh Ippen C, Schneider LS. Severity of cognitive impairment in Alzheimer's disease affects list learning using the California Verbal Learning Test (CVLT). *International Journal of Geriatric Psychiatry*. 1998; 13:544–549. [PubMed: 9733335]
- Ganguli M, Dodge HH, Shen C, DeKosdy ST. Mild cognitive impairment, amnesic type. *Neurology*. 2004; 63:115–121. [PubMed: 15249620]
- Goodglass, H.; Kaplan, E.; Barresi, B. Boston Diagnostic Aphasia Examination. Third Edition. Lippincott Williams & Wilkins; Philadelphia: 2001.
- Griffith HR, Netson KL, Harrell LE, Zamrini EY, Brockington JC, Marson DC. Amnesic mild cognitive impairment: Diagnostic outcomes and clinical prediction over a two-year time period. *Journal of the International Neuropsychological Society*. 2006; 12:166–175. [PubMed: 16573850]
- Grober E, Sliwinski M. Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*. 1991; 13:933–949. [PubMed: 1779032]
- Guarch J, Marcos T, Salameiro M, Gastó C, Blesa R. Mild cognitive impairment: A risk indicator of later dementia, or a preclinical phase of the disease? *International Journal of Geriatric Psychiatry*. 2008; 23:257–265. [PubMed: 17668419]
- Gurol ME, Irizarry MC, Smith EE, Raju S, Diaz-Arrastia R, Bottiglieri T, et al. Plasma B-amyloid and white matter lesions in AD, MCI, and cerebral amyloid angiopathy. *Neurology*. 2006; 66:23–29. [PubMed: 16401840]
- Hachinski VC, Iliff LD, Zilka E, Du Boulay GH, McAllister VL, Marshall J, et al. Cerebral blood flow in dementia. *Archives of Neurology*. 1975; 32:632–637. [PubMed: 1164215]
- Hamilton M. Development and rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*. 1967; 6:278–296. [PubMed: 6080235]
- Heaton, RK.; Chelune, GJ.; Talley, JL.; Kay, GG.; Curtiss, G. Wisconsin Card Sorting Test manual: Revised and expanded. Psychological Assessment Resources; Lutz, FL: 1993.
- Holtzer R, Goldin Y, Zimmerman M, Katz M, Buschke H, Lipton RB. Robust norms for selected neuropsychological tests in older adults. *Archives of Clinical Neuropsychology*. 2008; 23:531–541. [PubMed: 18572380]
- Howieson DB, Carlson NE, Milar Moore M, Wasserman D, Abendroth CD, Payne-Murphy J, Kaye JA. Trajectory of mild cognitive impairment onset. *Journal of the International Neuropsychological Society*. 14:192–198.
- Ivanoiu A, Adams S, Van der Linden M, Salmon E, Juillerat A-C, Mulligan R, Seron X. Memory evaluation with a new cued recall test in patients with mild cognitive impairment and Alzheimer's disease. *Journal of Neurology*. 2005; 252:47–55. [PubMed: 15654553]
- Jorm AF, Scott R, Jacomb PA. Assessment of cognitive decline in dementia and by informant questionnaire. *International Journal of Geriatric Psychiatry*. 1989; 4:35–9.

- Jurica, P.; Leitten, C.; Mattis, S. Dementia Rating Scale-2. Psychological Assessment Resources; Lutz, FL: 2001.
- Karrasch M, Sinerva E, Gronholm P, Rinne J, Laine M. CERAD test performances in amnesic mild cognitive impairment and Alzheimer's disease. *Acta Neurologica Scandinavica*. 2005; 111:172–179. [PubMed: 15691286]
- Kasai M, Megura K, Hashimoto R, Ishizaki J, Yamadori A, Mori E. Non-verbal learning is impaired in very mild Alzheimer's disease (CDR 0.5): Normative data from the learning version of the Rey-Osterrieth Complex Figure Test. *Psychiatry and Clinical Neuroscience*. 2006; 60:139–146.
- Kavé G, Heinik J. Neuropsychological evaluation of mild cognitive impairment: Three case reports. *The Clinical Neuropsychologist*. 2004; 18:362–372. [PubMed: 15739808]
- Lezak, MD.; Howieson, DB.; Loring, DW. Neuropsychological assessment. 4th ed.. Oxford University Press; New York: 2004.
- Marruff P, Collie A, Darby D, Weaver-Cargin J, Masters C, Currie J. Subtle memory decline over 12 months in mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*. 2004; 18:342–348. [PubMed: 15316183]
- McCaffrey, RJ.; Duff, K.; Westervelt, HJ. Practioner's Guide to Evaluating Change With Neuropsychological Assessment Instruments. Kluwer Academic/Plenum Publishers; New York: 2000.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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*. 1984; 34:939–944. [PubMed: 6610841]
- Modrego PJ. Predictors of conversion to dementia of probable Alzheimer type in patients with mild cognitive impairment. *Current Alzheimer Research*. 2006; 3:161–170. [PubMed: 16611017]
- Morris JC. Mild cognitive impairment is early-stage Alzheimer disease. *Archives of Neurology*. 2006; 63:15–16. [PubMed: 16401731]
- Morris JC. The Clinical Dementia Rating: Current version and scoring rules. *Neurology*. 1993; 43:2412–2414. [PubMed: 8232972]
- Morris JC, Heyman A, Mohs R, Hughes JP, van Belle G, Fillenbaum G. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989; 44:609–614.
- Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, et al. Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's & Dementia*. 2005; 1:55–56.
- Petersen RC. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*. 2004; 256:183–194. [PubMed: 15324362]
- Petersen RC, Negash S. Mild cognitive impairment: An overview. *CNS Spectrums*. 2007; 45:53.
- Petersen RC, Smith GE, Ivnik RT, Kokmen E, Tangalos EG. Memory function in very early Alzheimer's disease. *Neurology*. 1994; 44:867–872. [PubMed: 8190289]
- Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*. 1999; 56:303–308. [PubMed: 10190820]
- Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*. 2001; 56:1133–1142. [PubMed: 11342677]
- Petrella JR, Krishnan S, Slavin MJ, Tran-Thu TT, Murty L, Doraiswamy PM. Mild cognitive impairment: Evaluation with 4-T functional MR imaging. *Radiology*. 2006; 240:177–186. [PubMed: 16684919]
- Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. *Journal of Gerontology*. 1982; 37:323–329. [PubMed: 7069156]
- Piatt AL, Fields JA, Paolo AM, Troster AI. Action (verb naming) fluency as an executive function measure: Convergent and divergent evidence of validity. *Neuropsychologia*. 1999; 37:1499–1503. [PubMed: 10617270]

- Piatt AL, Fields JA, Paolo AM, Troster AI. Action verbal fluency normative data for the elderly. *Brain and Language*. 2004; 89:580–583. [PubMed: 15120548]
- Rabin LA, Borgos MJ, Saykin AJ, Wishart HA, Crane PK, Nutter-Upham KE, et al. Judgment in older adults: Development and psychometric evaluation of the Test of Practical Judgment (TOP-J). *Journal of Clinical and Experimental Neuropsychology*. 2007; 29:752–767. [PubMed: 17896200]
- Rabin LA, Saykin AJ, Wishart HA, Nutter-Upham KE, Flashman LA, Paré N, et al. The Memory and Aging Telephone Screen (MATS): Development and preliminary validation. *Alzheimer's & Dementia*. 2007; 3:109–121.
- Reitan, R.; Wolfson, D. The Halstead-Reitan Neuropsychological Test Battery: theory and clinical interpretation. Neuropsychology Press; Tucson, AZ: 1985.
- Ritchie LJ, Frerichs RJ, Tuokko H. Effective normative samples for the detection of cognitive impairment: A population-based validation study. *Neurology*. 2007; 56:37–42. [PubMed: 11148233]
- Salmon DP, Thomas RG, Pay MM, Booth A, Hofstetter CR, Thal LJ, Katzman R. Alzheimer's disease can be accurately diagnosed in very mildly impaired individuals. *Neurology*. 2002; 59:1022–1028. [PubMed: 12370456]
- Santulli, RB.; Saykin, AJ.; Rabin, LA.; Wishart, HA.; Flashman, LA.; Paré, N., et al. Differential sensitivity of cognitive complaints associated with amnesic MCI: Analysis of patient and informant reports. Alzheimer's Association International Conference on Prevention of Dementia; Washington, D.C.. 2005.
- Saxton J, Lopez OL, Ratcliff G, Dulberg C, Fried LP, Carlson MC, et al. Preclinical Alzheimer disease: Neuropsychological test performance 1.5 to 8 years prior to onset. *Neurology*. 2004; 63:2341–2347. [PubMed: 15623697]
- Saykin AJ, Wishart HA, Rabin LA, Flashman LA, Mamourian AC, Santulli RB. Cholinergic enhancement of frontal lobe activity in Mild Cognitive Impairment. *Brain*. 2004; 127:1574–83. [PubMed: 15140813]
- Saykin, AJ. Dartmouth Medical School. NBFADL-63 item version. 1992. 1992. Neurobehavioral function and activities of daily living rating scale. available from author
- Saykin AJ, Janssen RS, Sprehn GC, Kaplan JE, Spira TJ, O'Connor B. Longitudinal evaluation of neuropsychological function in homosexual men with HIV infection 18-month follow-up. *Journal of Neuropsychiatry and Clinical Neurosciences*. 1991; 3:286–298. [PubMed: 1821245]
- Saykin AJ, Wishart HA, Rabin LA, Santulli RB, Flashman LA, West JD, et al. Older adults with cognitive complaints show brain atrophy similar to that of amnesic MCI. *Neurology*. 2006; 67:834–862. [PubMed: 16966547]
- Schmidt, M. Rey Auditory and Verbal Learning Test: A handbook. Western Psychological Services; Los Angeles: 1996.
- Shankle WR, Romney AK, Hara J, Fortier D, Dick MB, Chen JM, Chan T, Sun X. Methods to improve the detection of mild cognitive impairment. *Proceedings of the National Academy of Sciences*. 2005; 102:4919–4924.
- Small BJ, Fratiglioni L, Viitanen M, Winblad B, Bäckman L. The course of cognitive impairment in preclinical Alzheimer's disease: 3- and 6-year follow-up of a population-based sample. *Archives of Neurology*. 2000; 57:839–844. [PubMed: 10867781]
- Small GW. Revised ischemic score for diagnosing multi-infarct dementia. *Journal of Clinical Psychiatry*. 1985; 46:514–517. [PubMed: 4066617]
- Spitznagel MB, Tremont G, Brown LB, Gunstad J. Cognitive reserve and the relationship between depressive symptoms and awareness of deficits in dementia. *Journal of Neuropsychiatry and Clinical Neurosciences*. 2006; 18:186–190. [PubMed: 16720795]
- Spreen, O.; Strauss, E. A compendium of neuropsychological tests: Administration, norms, and commentary. 2nd ed.. Oxford University Press; London: 1998.
- Squire LR, Wetzel CD, Slater PC. Memory complaints after electroconvulsive therapy: Assessment with a new self-rating instrument. *Biological Psychiatry*. 1979; 14:791–801. [PubMed: 497304]
- Stern Y. What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*. 2002; 8:448–60. [PubMed: 11939702]

- Thal, LJ. Alzheimer's Disease Cooperative Study (ADCS) overview. In: Vellas, B.; Fitten, LJ.; Winblad, B.; Feldman, H.; Grundman, M.; Giacobini, E., editors. Research and Practice in Alzheimer's Disease. Vol. Vol. 10. Serdi; Paris: 2005. p. 25-33.
- Tierney MC, Szalai JP, Dunn E, Geslani D, McDowell I. Prediction of probable Alzheimer disease in patients with symptoms suggestive of memory impairment. Archives of Family Medicine. 2000; 9:527–532. [PubMed: 10862215]
- Tierney MC, Yao C, Kiss A, McDowell I. Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. Neurology. 2005; 64:1853–1859. [PubMed: 15955933]
- Troyer AK, Murphy KJ, Anderson ND, Hayman-Abello BA, Craik FIM, Moscovitch M. Item and associative memory in amnesic mild cognitive impairment: Performance on standardized memory tests. Neuropsychology. 2008; 22:10–16. [PubMed: 18211151]
- Wechsler, D. Wechsler Adult Intelligence Scale. Third Edition. The Psychological Corporation; San Antonio, TX: 1997.
- Wechsler, D. Wechsler Memory Scale-Third Edition, WMS-III administration and scoring manual. The Psychological Corporation; San Antonio, TX: 1997.
- Wicklund AH, Johnson N, Rademaker A, Weitner BB, Weintraub S. Word list versus story memory in Alzheimer disease and frontotemporal dementia. Alzheimer Disease and Associated Disorders. 2006; 20:86–92. [PubMed: 16772743]
- Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund L-O, et al. Mild cognitive impairment - beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. Journal of Internal Medicine. 2004; 256:240–246. [PubMed: 15324367]
- Yaari R, Fleisher AS, Gamst AC, Bagwell VP, Thal LJ. Utility of the telephone interview for cognitive status for enrollment in clinical trials. Alzheimer's & Dementia. 2006; 2:104–109.
- Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO. Development and validation of a geriatric depression scale: A preliminary report. Journal of Psychiatric Research. 1982; 17:37–49. [PubMed: 7183759]

Table 1

Participant Demographics and Scores on Global Cognitive Tests and Performance Indicators

	HC n = 51	MCI n = 38	Effect Size	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t</i>	<i>d</i>
Sex (Male / Female)	16 / 35	18 / 20 ^a		
Age	70.63 (5.06)	71.58 (6.21)	-0.80	-0.17
Education	16.82 (2.60)	16.55 (3.16)	0.44	0.09
MMSE (/ 30) 100%	29.18 (0.99)	26.76 (2.19)	6.98	1.42**
DRS-2 (/ 144) 100%	140.90 (2.20)	136.74 (4.35)	5.91	1.21**
VOCSS 24.3%	13.96 (1.81)	13.32 (2.71)	1.34	0.28
VR1SS 100%	11.92 (2.55)	8.29 (3.24)	5.92	1.25**
VRDSS 100%	12.82 (3.06)	8.63 (2.49)	6.91	1.50**
VRRECSS 100%	12.84 (2.05)	9.74 (2.44)	6.52	1.38**
DRS-MSS 99.9%	10.82 (1.95)	7.66 (3.31)	5.65	1.16**
CVLT-II learning (/ 80) 100%	48.71 (7.79)	32.79 (6.88)	10.02	2.17**
CVLT-II short delay (/ 16) 100%	10.96 (2.60)	5.55 (2.54)	9.81	2.10**
CVLT-II long delay (/ 16) 100%	11.67 (2.67)	5.63 (2.61)	10.65	2.30**
LM I (/ 75) 100%	47.47 (7.35)	33.50 (9.15)	7.99	1.68**
LM II (/ 50) 100%	31.25 (6.14)	17.92 (8.07)	8.51	1.86**
LMREC (/ 30) 100%	28.12 (1.48)	23.59 (3.03)	5.79	2.00**
LM II-A (/ 25) 100%	14.94 (3.61)	7.53 (4.54)	8.30	1.81**
MATS learning (/ 30) 100%	20.70 (3.71)	16.47 (3.65)	5.00	1.15**
MATS delay (/ 10) 100%	6.47 (1.73)	3.35 (1.80)	8.40	1.77**
MMSE W-Missed (/ 3) 100%	0.59 (0.90)	2.13 (1.04)	-7.48	-1.58**
INFO (/ 28) 78.1%	23.10 (3.07)	20.58 (5.02)	2.93	0.62*

HC = healthy control; MCI = mild cognitive impairment; MMSE = Mini-Mental State Examination, total score; DRS-2 = Dementia Rating Scale, total score; VOCSS = WAIS-III Vocabulary age scaled score; VR1SS = WMS-III Visual Reproduction Immediate Recall age scaled score; VRDSS = WMS-III Visual Reproduction Delay Recall age scaled score; VRRECSS = WMS-III Visual Reproduction Recognition age scaled score; DRSMSS = Dementia Rating Scale-2 Memory age scaled score; CVLT-II learning = California Verbal Learning Test-II total learning raw score; CVLT-II short delay = California Verbal Learning Test-II short delay free recall raw score; CVLT-II long delay = California Verbal Learning Test-II long delay free recall raw score; LM I = WMS-III Logical Memory immediate recall raw score; LM II = WMS-III Logical Memory delay raw score; LMREC = WMS-III Logical Memory Recognition raw score; LM II-A = Logical Memory Story A delay recall raw score; MATS learning = Memory and Aging Telephone Screen total learning raw score; MATS delay = Memory and Aging Telephone Screen long delay raw score; MMSE W-missed = Mini-Mental State Examination, number of words missed of three; INFO = WAIS-III Information subtest raw score

^a Chi square value not significant: ($\chi^2 = 2.36, df = 1, p = .124$)

* $p < .05$

** $p < .001$

Table 2

Sensitivity and Specificity of Memory Measures Ordered by Overall Accuracy

Memory indices	Sensitivity (%)	Specificity (%)	Accurate classification (%)
CVLT-II learning	90.2	84.2	87.6
LMREC	92.2	83.8	86.6
CVLT-II long delay	88.2	84.2	86.5
LM II-A	92.2	76.3	85.4
MATS delay	86.0	82.4	84.4
CVLT-II short delay	84.3	78.9	82.0
LM II	86.3	73.7	80.9
MMSE W-Missed	84.3	73.7	79.8
MATS learning	83.7	70.6	77.9
LM I	86.3	65.8	77.5
INFO	80.4	52.6	68.5

CVLT-II learning = California Verbal Learning Test-II total learning raw score; LMREC = WMS-III Logical Memory recognition raw score; CVLT-II long delay = California Verbal Learning Test-II long delay free recall raw score; LM II-A = Logical Memory Story A delay recall raw score; MATS delay = Memory and Aging Telephone Screen long delay raw score; CVLT-II short delay = California Verbal Learning Test-II short delay free recall raw score; LM II = WMS-III Logical Memory delay raw score; MMSE W-missed = Mini-Mental State Examination, number of words missed of three; MATS learning = Memory and Aging Telephone Screen total learning raw score; LM I = WMS-III Logical Memory immediate recall raw score; INFO = WAIS-III Information subtest raw score

Table 3

Summary of Logistic Regression Analysis Predicting MCI

Memory indices	<i>B</i>	<i>SE</i>	Odds ratio	Wald Statistic
CVLT-II learning	−0.31	0.07	0.74	22.44**
LMREC	−1.05	0.23	0.35	20.78**
CVLT-II long delay	−0.81	1.75	0.45	21.42**
LM II-A	−0.41	0.09	0.66	20.68**
MATS delay	−1.17	0.28	0.31	18.09**
CVLT-II delay	−0.73	0.16	0.48	22.31**
LM II	−0.28	0.06	0.76	20.28**
MMSE W-Missed	1.32	0.26	3.75	25.32**
MATS learning	−0.37	0.10	0.69	13.86**
LM I	−0.20	0.04	0.82	23.03**
INFO	−0.24	0.08	0.79	9.69*

CVLT-II learning = California Verbal Learning Test-II total learning raw score; LMREC = WMS-III Logical Memory recognition raw score; CVLT-II long delay = California Verbal Learning Test-II long delay free recall raw score; LM II-A = Logical Memory Story A delay recall raw score; MATS delay = Memory and Aging Telephone Screen long delay raw score; CVLT-II short delay = California Verbal Learning Test-II short delay free recall raw score; LM II = WMS-III Logical Memory delay raw score; MMSE W-missed = Mini-Mental State Examination, number of words missed of three; MATS learning = Memory and Aging Telephone Screen total learning raw score; LM I = WMS-III Logical Memory immediate recall raw score; INFO = WAIS-III Information subtest raw score

*
 $p < .01$

**
 $p < .001$

Table 4

Participant Demographics and Scores on Performance Indicators for AD Subgroup Analysis

	AD Group n = 9	Non-AD Group n = 23	Effect Size	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t</i>	<i>d</i>
Sex (Male / Female)	4 / 5	10 / 13 ^a		
Age	69.78 (7.66)	72.39 (5.54)	1.08	−0.39
Education	16.78 (2.73)	16.30 (3.27)	−0.38	0.07
CVLT-II total learning	31.44 (7.68)	33.13 (7.23)	0.58	−0.23
CVLT-II short delay	3.56 (2.65)	6.09 (2.41)	2.60	−1.00*
CVLT-II long delay	3.89 (3.26)	6.43 (2.30)	2.50	−0.90*
LM I	28.67 (6.34)	35.74 (9.24)	2.10	−0.89*
LM II	10.11 (6.49)	21.48 (6.86)	4.27	−1.70**
LMREC	20.50 (3.07)	24.30 (2.62)	3.39	−1.33**
LM II-A	2.89 (3.14)	9.48 (3.95)	4.47	−1.85**
MATS Total	17.11 (5.26)	16.43 (3.03)	−0.46	0.16
MATS delay	1.89 (1.62)	3.96 (1.58)	3.31	−1.29**
MMSE W-Missed	2.22 (1.09)	2.17 (0.98)	−0.12	0.05
INFO	20.11 (3.82)	20.09 (5.86)	−0.11	0.00

AD = Alzheimer's disease; CVLT-II learning = California Verbal Learning Test-II total learning raw score; CVLT-II short delay = California Verbal Learning Test-II short delay free recall raw score; CVLT-II long delay = California Verbal Learning Test-II long delay free recall raw score; LM I = WMS-III Logical Memory immediate recall raw score; LM II = WMS-III Logical Memory delay raw score; LMREC = WMS-III Logical Memory recognition raw score; LM II-A = Logical Memory Story A delay recall raw score; MATS learning = Memory and Aging Telephone Screen total learning raw score; MATS delay = Memory and Aging Telephone Screen long delay raw score; MMSE W-missed = Mini-Mental State Examination, number of words missed of three; INFO = WAIS-III Information subtest raw score

^aChi square value not significant: ($\chi^2 = .002$, $df = 1$, $p = .96$)

*
 $p < .05$

**
 $p < .001$

Table 5

Sensitivity and Specificity of Memory Measures for Predicting Conversion from MCI to AD Ordered by Overall Accuracy

Memory indices	Sensitivity (%)	Specificity (%)	Accurate classification (%)
LMREC	95.7	62.5	87.1
CVLT-II long delay	100.0	44.4	84.4
LM II-A	87.0	66.7	81.3
MATS delay	91.3	44.4	78.1
CVLT-II short delay	91.3	44.4	78.1
LM I	91.3	33.3	75.0
LM II	82.6	55.6	75.0
CVLT-II learning	95.7	11.1	71.9
MMSE W-Missed	100.0	00.0	71.9
INFO	100.0	00.0	71.9
MATS learning	100.0	00.0	71.9

CVLT-II long delay = California Verbal Learning Test-II long delay free recall raw score; LMREC = WMS-III Logical Memory recognition raw score; LM I = WMS-III Logical Memory immediate recall raw score; LM II-A = Logical Memory Story A Delay Recall raw score; MATS delay = Memory and Aging Telephone Screen long delay raw score; CVLT-II short delay = California Verbal Learning Test-II short delay free recall raw score; LM II = WMS-III Logical Memory delay raw score; CVLT-II learning = California Verbal Learning Test-II total learning raw score; MMSE W-missed = Mini-Mental State Examination, number of words missed of three; INFO = WAIS-III Information subtest raw score; MATS learning = Memory and Aging Telephone Screen total learning raw score

Note. The average duration of follow-up was 2.97 yrs, SD = 1.16 (range = 1 – 4 years).

Table 6

Summary of Logistic Regression Analysis Predicting Progression to AD

Memory indices	<i>B</i>	<i>SE</i>	Odds ratio	Wald Statistic
LMREC	−0.63	0.25	0.53	6.41 *
CVLT-II long delay	−0.82	0.34	0.44	5.80 *
LM II-A	−0.43	0.16	0.65	7.19 **
MATS delay	−0.82	0.39	0.44	5.80 *
CVLT-II short delay	−0.53	0.24	0.59	5.06 *
LM I	−0.18	0.08	0.84	4.88 *
LM II	−0.23	0.09	0.79	7.35 *
CVLT-II learning	−0.07	0.06	0.93	1.13
MMSE W-Missed	0.17	0.44	1.18	0.14
INFO	−0.01	0.12	0.99	0.00
MATS learning	0.01	0.11	1.02	0.02

CVLT-II long delay = California Verbal Learning Test-II long delay free recall raw score; LMREC = WMS-III Logical Memory recognition raw score; LM I = WMS-III Logical Memory immediate recall raw score; LM II-A = Logical Memory Story A delay recall raw score; MATS delay = Memory and Aging Telephone Screen long delay raw score; CVLT-II short delay = California Verbal Learning Test-II short delay free recall raw score; LM II = WMS-III Logical Memory delay raw score; CVLT-II learning = California Verbal Learning Test-II total learning raw score; MMSE W-missed = Mini-Mental State Examination, number of words missed of three; INFO = WAIS-III Information subtest raw score; MATS learning = Memory and Aging Telephone Screen total learning raw score

*
 $p < .05$

**
 $p < .01$