Journal of Clinical and Experimental Neuropsychology, 28:1408–1422, 2006

ISSN: 1380-3395
DOI: 10.1080/13803390500409583

Psychology Press
Taylor & Francis Group

Memory Evaluation in Mild Cognitive Impairment using Recall and Recognition Tests

ILANA J. BENNETT,[1] EDWARD J. GOLOB,[1,2]
ELIZABETH S. PARKER,[1] AND ARNOLD STARR[1]

[1]Department of Neurology, University of California, Irvine, USA
[2]Institute for Brain Aging and Dementia, University of California, Irvine, USA

*Amnestic mild cognitive impairment (MCI) is a selective episodic memory deficit that often indicates early Alzheimer's disease. Episodic memory function in MCI is typically defined by deficits in free recall, but can also be tested using recognition procedures. To assess both recall and recognition in MCI, MCI (*n *= 21) and older comparison (*n *= 30) groups completed the USC-Repeatable Episodic Memory Test. Subjects memorized two verbally presented 15-item lists. One list was used for three free recall trials, immediately followed by yes/no recognition. The second list was used for three-alternative forced-choice recognition. Relative to the comparison group, MCI had significantly fewer hits and more false alarms in yes/no recognition, and were less accurate in forced-choice recognition. Signal detection analysis showed that group differences were not due to response bias. Discriminant function analysis showed that yes/no recognition was a better predictor of group membership than free recall or forced-choice measures. MCI subjects recalled fewer items than comparison subjects, with no group differences in repetitions, intrusions, serial position effects, or measures of recall strategy (subjective organization, recall consistency). Performance deficits on free recall and recognition in MCI suggest a combination of both tests may be useful for defining episodic memory impairment associated with MCI and early Alzheimer's disease.*

Introduction

Mild cognitive impairment describes the transition state between healthy aging and early dementia that is accompanied by declines in one or multiple cognitive domains (Flicker, Ferris & Reisberg, 1991; Petersen et al., 2001). The amnestic subtype of mild cognitive impairment (MCI) describes older individuals with an episodic memory deficit more severe than is expected for normal aging (Fercker et al., 1991; Pafewen, 2003; Smith et al., 1996). Amnestic MCI patients do not meet the criteria for dementia because other measures of cognitive function are within normal limits, and activities of daily living are unaffected (APA, 1994; Petersen, 2004). However, individuals with MCI have approximately a six-fold increased risk of developing Alzheimer's disease relative to older people without episodic memory impairments (Petersen et al., 1999). In addition, β-amyloid plaques, which are characteristic of Alzheimer's disease, have been reported in postmortem studies

Received 6 February 2005; accepted 6 October 2005.
This work was supported by NIA grant AG19681. The authors wish to thank Carl Cotman for support and the UC Irvine Alzheimer's Disease Research Center staff. We also thank Rie Irimajiri for valuable discussions concerning these experiments.
Address correspondence to Edward J. Golob, Ph.D. Department of Psychology, 3067 Percival Stern Hall, Tulane University, New Orleans, LA 70118. E-mail: egolob@tulane.edu

of MCI (Morris et al., 2001; Pirce & Morris, 1999). Taken together, these observations suggest MCI is often an indicator of early Alzheimer's disease.

Episodic memory functions in MCI are typically assessed with tests of recall (Collie & Maruff, 2000; Morris et al., 2001). In studies of normal aging, recall performance shows a larger age-related decline compared to recognition (Craik & McDowd, 1987; Moscovitch & Winoau, 1990; Parker, Landau, Whipple & Schwartz, 2004). Age-related changes in recall may make it difficult to distinguish between early Alzheimer's disease and normal aging because low performing normals may be considered MCI. Recall performance also declines in a variety of neurological disorders besides Alzheimer's disease, such as depression, schizophrenia, Huntington's disease, and Parkinson's disease (Zakzanis, Leach & Kaplan, 1999). Thus, as compared to tests of recall, tests using a recognition format may be helpful for the early detection of Alzheimer's disease because recognition scores are less affected by normal aging, and may have greater specificity.

Recall and recognition tests can evaluate memory differently by manipulating the amount of information provided during memory retrieval (Burke & Light, 1981). In free recall tests subjects must produce studied items without any external retrieval cues. In contrast, recognition tests present stimuli that serve as cues for the retrieval of studied items. Yes/no and forced choice recognition are two commonly used recognition formats, with each format providing different types of recognition cues. In yes/no recognition, previously studied items (targets) and new items (distractors) are randomly presented one at a time, and subjects respond whether the item was ("yes") or was not ("no") in the study list. Forced-choice recognition tests present at least two items simultaneously, one of which is the target, and subjects attempt to identify the target. Previous studies have found no significant differences between performance on yes/no and forced-choice recognition tests in younger subjects (Bastin & Van der Linden, 2003; Khoe, Kroll, Yonelinas, Dobbins & Knight, 2000; cf. Kroll, Yonelinas, Dobbins & Frederick, 2002). However, a significant effect of aging has been reported, with larger age-related declines in yes/no relative to forced-choice recognition (Bastin & Van der Linden, 2003). Performance declines are seen in yes/no testing in both MCI (Golob, Johnson & Starr, 2002; Wang & Zhou, 2002) and Alzheimer's disease (Backman, Small & Fratiglioni, 2001; Morris et al., 1989) compared to healthy older controls. Forced-choice tests also show declines in both MCI (Machulda et al., 2003) and Alzheimer's disease (Christensen, Koplman, Stanhope, Lorentz & Owen, 1998; Hodges & Patterson, 1995). It is unclear if a particular test format (yes/no vs. forced-choice) is superior for defining differences in MCI or Alzheimer's disease because previous studies have not directly compared the different test formats.

The present study used the University of Southern California-Repeatable Episodic Memory Test (USC-REMT) to assess episodic memory performance in a relatively small sample of amnestic MCI (n = 21) using a combination of free recall, yes/no recognition, and forced-choice recognition. The USC-REMT is a recently published brief (~10 minute) screening test with several different versions that allow for repeated testing. Previous studies indicate that the USC-REMT can detect age-related changes in episodic memory function (Parker, Eaton, Whipple, Heseltine & Brigde, 1995; Parker et al., 2004). Recall was assessed using standard measures (number correct, repetitions, intrusions) in conjunction with measures of organizational strategy (subjective organization, recall consistency) and recall as a function of serial position. Recognition measures included hits, false alarms, and proportion of hits minus proportion of false alarms (Pr) in the yes/no test; and percent correct in forced-choice recognition. Signal detection methods were used to estimate each subject's ability to distinguish targets from distractors independent of response bias.

Methods

Subjects

Twenty-one MCI (15 M: 6 F) and 30 older comparison subjects (15 M: 15 F) participated in the study. The University of California, Irvine Institutional Review Board approved the experimental procedures, and all participants gave informed consent.

Sixteen MCI subjects were recruited from the University of California, Irvine Alzheimer's Disease Research Center (ADRC) where they received annual evaluations that included neuropsychological testing, neurological and physical examinations, neuroimaging, and family interviews. The neuropsychological test battery included several domains of cognitive function. The Mini-Mental State Examination (Folstein, Folstein & McHugh, 1975) and Activities of Daily Living scale (Galasko et al., 1997) were used to screen for dementia. Episodic memory was assessed with the WMS-III Logical Memory subtest (Wechsler, 1997) and CERAD Word List Learning Task (Morris et al., 1989). Language ability was measured with the 30-item version of the Boston Naming test (Kaplan, Snodgrass & Weintraub, 1983), CERAD Animal Naming (Morris et al., 1989), and Controlled Oral Word Association letter fluency (Spreen & Benton, 1977). Trailmaking Test A and B (Reitan, 1958) were used to test executive function, and visual-spatial skills were evaluated with WAIS-III Block Design subtest (Wechsler, 1997) and CERAD Constructional Praxis test (Morris et al., 1989).

The remaining five MCI subjects were recruited from outside the ADRC cohort. Three were recruited from a private clinical practice, one from a clinical drug study, and one from a long-term health study. MCI subjects were diagnosed based on criteria for the amnestic subtype: memory complaint, age-related memory impairment, preserved cognitive function in nonmemory domains, activities of daily living unaffected, and not demented (Petersen et al., 2001; Petersen, 2004). The USC-REMT was not used for diagnostic purposes.

Comparison subjects were selected from an existing data set (Parker et al., 2004) to match the MCI group for age and education level. Comparison subjects were living independently in the community and presented with no neurological problems. Neuropsychological evaluations were not conducted in comparison subjects. However, results indicate that levels of accuracy for recall and recognition were comparable to previous studies of episodic memory in aging.

All MCI subjects received a battery of neuropsychological tests examining episodic memory and other cognitive domains. The battery of neuropsychological tests and the present study were conducted on separate days. In contrast, the comparison group was only given the USC-REMT. To characterize the results from MCI on tests other than the USC-REMT, results from a group of normative subjects are shown in Table 2. Comparisons were between MCI ($n = 16$) and a normative sample of age and education matched older subjects that received the same test battery as part of a Successful Aging Program at the UC Irvine ADRC. Results indicated that individual MCI subjects performed within the normal range ($\pm$ 1.5 SD) on all cognitive domains except episodic memory, which was typically $\geq$ 1.5 SD below the mean of age-matched norms. There was also a small group difference in picture naming, as determined using the Boston Naming Test. We emphasize that the normative data in Table 2 are presented solely for the purpose of characterizing performance in MCI in several cognitive domains relative to healthy older subjects. The main findings of the present study compared performance on the USC-REMT between MCI and a group of comparison subjects that did not receive additional neuropsychological testing.

Eleven of the 21 MCI subjects were taking cholinesterase inhibitors at the time of testing. Free recall correct showed a significant effect of medication, with nonmedicated MCI (15.8 ± 2.0) recalling more that medicated MCI (11.7 ± 2.0) ($p < .001$), suggesting that subjects with more advanced stages of amnestic MCI were taking medication. Medication did not affect other USC-REMT or neuropsychological test measures, expect for a slight difference on the CERAD constructional praxis test ($p < .02$), showing better visual-spatial skills in nonmedicated versus medicated MCI.

Test Description and Procedures

The University of California-Repeatable Episodic Memory Test (USC-REMT) is a verbal test of free recall, yes/no recognition, and three-alternative forced-choice recognition memory (Parker et al., 2004, 1995). There were seven different lists having 45 semantically unrelated, high frequency nouns (15 target items, 30 distractors) that are read to subjects. One list was used for free recall and yes/no recognition; a second list was used for forced-choice recognition. See Parker et al. (2004) for additional test list and administration information.

The free recall test had three study-test trials. For the study phase of each trial, subjects were instructed to memorize 15 target words that were orally presented by the experimenter at a rate of one item every 2 sec. During the test phase, the subject tried to recall as many items as possible in any order. The number of correctly recalled target items (correct), target items that were previously recalled on that trial (repetitions), and words that were not on the study list (intrusions) were recorded for each trial. Target items were presented in a different order for each trial to assess subjective organization (described below).

The yes/no recognition test immediately followed the third recall trial. A list was presented that included the same 15 target items as the free recall task plus 30 unrelated distractor items presented in random order. The experimenter orally presented the words one at a time at a pace set by the subject's response rate. Subjects were instructed to respond "yes" if the item was previously studied on the free recall list, and "no" if the item was new. Correct "yes" responses to targets were considered hits and incorrect "yes" responses to distractors were labeled false alarms.

In the forced-choice recognition test there was a single study phase, with the experimenter orally presenting the 15 targets at a rate of one item every 2 sec. There was no recall test following the list presentation. During the test phase the experimenter read groups of three words (one target item and two distractors). Subjects were told that in each trial one of the three words was a member of the studied list, and they were to indicate which word was previously studied. The number of items that were correctly identified as having been presented on the study list was recorded.

The order of test presentation (free recall and yes/no, forced-choice) was counterbalanced across subjects. Two of the seven alternative study lists were randomly selected for each subject, and the order of test presentation was counterbalanced across subjects within each group. Responses during each of the tests were self-paced, and there was no delay between list presentation and recall or recognition of test items. For both recognition tests, subjects were instructed to respond to each item even if they had to guess. The entire testing period lasted ~10 minutes.

Statistical Analysis

Free recall measures included the number of correctly recalled items, repetitions, intrusions, recall consistency, subjective organization, and serial position. Recall consistency indexes

the extent that the same items were recalled across trials, regardless of order (Bousfield & Bousfield, 1966). Recall consistency was defined as: ((T1&T2) + (T2&T3))/T1 + T2. Abbreviations indicate the number of items correctly recalled on both trials 1 and 2 (T1&T2), the number of items correctly recalled on both trials 2 and 3 (T2&T3), and the total number of items recalled on trials 1 and 2 (T1 + T2). Scores range from 0–100%, with larger scores indicating more of the same items were recalled across trials. Subjective organization measures the number of pairs of items that are recalled in succession on consecutive trials (Sternberg & Tulving, 1977, equation 5). For example, if serial recall of the item "letter" was followed by "future" on the first trial, then serial recall of "letter" followed by "future", or "future" followed by "letter", on the next trial would count as one pair. Subjective organization was calculated using the formula: $((T_n\&T_{n+1}\text{ pairs}) - 2C(C\text{-}1))/(T_n)(T_{n+1})$. $T_n\&T_{n+1}$ pairs indicates the number of pairs recalled on successive trials, and C is the number of items recalled on both trials. Subjective organization was calculated for trials 1 and 2, and between trials 2 and 3. To assess group differences in the effects of serial position, the percent of correctly recalled items from the primacy (items 1–5), middle (item 6–10), or recency (items 11–15) positions of the presented target list were averaged across the three recall trials.

Yes/no test performance was analyzed using percent of hits (# of "yes" responses to targets divided by 15), percent of false alarms (# of "yes" responses to distractors divided by 30), and Pr (proportion of hits minus proportion of false alarms) in the yes/no test. The percent correct was recorded in the forced-choice test (15 possible).

In the yes/no test hits and false alarms cannot be assessed independently because they are influenced by the subject's response bias. For example, a tendency to respond "yes" would increase both hits and false alarms, while a tendency to respond "no" would result in fewer hits and false alarms. Signal detection theory was used to separately estimate the ability to distinguish between target and distractor items, sensitivity (d'), from response bias (c) (Green & Swets, 1966). Sensitivity and response bias were calculated from z-scores of hits and false alarms. A correction was included if the proportion of hits was equal to 1.0 (1–1/2n) and if the proportion of false alarms was equal to 0.0 (1/2n) (Macmillan & Creelman, 1991). For comparison with the yes/no test, d' was also measured in the forced-choice test.

Recall data were analyzed using analyses of variance (ANOVA) with factors of group (MCI, comparison), trial (one, two, three), and serial position (primacy, middle, recency). The Greenhouse-Geisser correction was used for repeated measures when appropriate. Independent sample *t* tests were also used to analyze group differences in repetitions, intrusions, recall consistency and subjective organization. Group differences in yes/no (hits, false alarms, Pr, d' and c) and forced-choice (correct, d') recognition measures were analyzed using independent sample *t* tests. Degrees of freedom prior to Greenhouse-Geisser correction are reported, and p values $< .05$ were considered significant.

Stepwise discriminant function analyses were used to quantify the ability of selected measures to classify individuals as belonging to either the MCI or comparison group. Subject classification used a jacknife procedure, where prediction of a given subject's status (MCI or comparison) was based upon a model that did not include that subject. Effect sizes in ANOVA were calculated using partial eta squared (η^2) (Tabachnick & Fidell, 1996).

Results

Demographics and Neuropsychological Test Battery

Demographic information for amnestic MCI and older comparison subjects is presented in Table 1. Neuropsychological results from a subset of MCI subjects that received the same

Table 1
Demographic information

	MCI	Comparison
n	21	30
Age	76.8 ± 4.9	78.9 ± 5.1
Education	15.2 ± 2.1	16.2 ± 2.9
Male/Female	15/6	15/15

Note. Mean ± SD.

battery of tests are shown in Table 2. Because comparison subjects did not receive a neuropsychological test battery, results from a normative group of older subjects enrolled in the Successful Aging Program at the UC Irvine ADRC are also shown in Table 2. Normative data are shown in Table 2 for the purpose of providing a reference to better appreciate performance levels in the MCI subjects. MCI subjects had significantly lower scores on all episodic memory tests relative to normative subjects ($p < .01$). There were no significant differences on the other neuropsychological tests, except for the Boston Naming Test ($p < .03$).

Free Recall

A 2 (group) × 3 (trial) ANOVA was used to analyze the number of correctly recalled items. There were significant main effects of group ($F_{(1, 49)} = 27.3$; $p < .001$), with comparison subjects recalling more items than MCI subjects (Figure 1A), and trial ($F_{(2, 98)} = 88.0$; $p < .001$), indicating greater recall across trials. There was also a significant group x trial interaction ($F_{(2, 98)} = 6.5$; $p < .01$), with comparison subjects having larger increases in the number of correctly recalled items across trials compared to MCI subjects (Figure 1B). Post hoc paired-sample *t* tests in comparison subjects showed significant increases in correctly recalled items between trials 1 and 2 ($t_{(29)} = –8.5$; $p < 0.001$), and trials 2 and 3 ($t_{(29)} = –5.0$; $p < 0.001$). In contrast, MCI subjects had significant increases between trials 1 and 2 ($t_{(20)} = –6.2$; $p < 0.001$), but not between trials 2 and 3. Independent sample *t* tests revealed no significant group differences for repetition or intrusion measures (Figure 1A).

Serial position was analyzed with a 2 (group) × 3 (serial position: primacy, middle, recency) ANOVA. A significant main effect of serial position ($F_{(2,98)} = 28.8$; $p < .001$) showed that both groups recalled more items from the end of the list (47.1 ± 2.1%) compared to the beginning (28.6 ± 1.9%) or middle (24.2 ± 1.6%) positions. Recall consistency and subjective organization were not significantly different between groups.

Recognition (Yes/no and Forced-choice)

Yes/no recognition had significant group differences in the percent of hits ($t_{(49)} = 3.9$; $p < .001$) and false alarms ($t_{(49)} = –4.7$; $p < .001$) (Figure 2A). MCI subjects had fewer hits and more false alarms relative to comparison subjects. The proportion of hits minus the proportion of false alarms (Pr) was calculated, and also showed significant group differences ($t_{(49)} = 8.7$; $p < .001$), with more accurate performance in the comparison group.

For the percent of correct responses in forced-choice recognition there was a significant group difference ($t_{(49)} = 4.9$; $p < .001$), with greater accuracy in comparison versus MCI subjects (Figure 2A).

Table 2
Neuropsychological test results

	MCI (n = 16)	Normative (n = 30)	p value
Cognitive status			
Mini-Mental State Examination	27.9 ± 1.9	28.7 ± 1.1	ns
Episodic memory			
CERAD Word List			
Immediate Recall Trials 1–3 (total)	13.8 ± 4.1	23.8 ± 3.1	< .01
5-min Delayed Recall	2.1 ± 1.8	8.1 ± 1.8	< .01
30-min Delayed Recall	1.1 ± 1.2	7.3 ± 2.2	< .01
5-min Delayed Recognition	17.3 ± 2.0	19.7 ± 0.6	< .01
30-min Delayed Recognition	15.8 ± 2.4	19.5 ± 0.8	< .01
WMS-III			
Logical Memory 1	23.6 ± 11.5	44.1 ± 8.8	< .01
Logical Memory 2	7.4 ± 8.5	28.1 ± 8.8	< .01
Language			
Letter Fluency	42.6 ± 10.1	44.3 ± 10.2	ns
Boston Naming Test (30-item)[a]	26.1 ± 3.1	28.3 ± 1.8	< .03
CERADAnimal Naming	17.5 ± 1.5	20.7 ± 4.9	ns
Executive function			
Trail Making Test A (sec)	40.6 ± 14.6	39.6 ± 14.3	ns
Trail Making Test B (sec)	120.1 ± 41.7	95.7 ± 46.0	ns
Visuo-spatial			
WAIS-III Block Design[b]	11.4 ± 2.7	12.4 ± 2.4	ns
CERAD Constructional Praxis	10.2 ± 1.0	10.4 ± 1.0	ns

Notes: Neuropsychological test results from a subset of MCI subjects in this study compared to a normative group (n = 30) of older subjects. Both groups completed the same neuropsychological test battery at the University of California, Irvine Alzheimer's disease Reserach Center. Normative subjects were matched to the MCI group for age (77.0 ± 5.3 years) and education level (15.4 ± 5.1 years). Normative subjects did not complete the USC-REMT, and are only included here as a comparison group for the MCI subjects. All scores are given as mean ± SD. Independent sample t test were used to show significant group differences ($p < .05$). ns = not significant.

[a]= Scores for one MCI subject were not included due to fatigue during testing.

[b]= Age-adjusted scaled scores.

Signal Detection Theory Analysis (Yes/no and Forced-choice)

Signal detection theory methods were used to define accuracy in distinguishing targets from distractor items (d') independently of response bias (c). Yes/no recognition showed significant group differences for d' ($t_{(49)} = -8.5$; $p < .001$), but not for response criterion (c). In forced-choice recognition there was a significant group difference for d' ($t_{(49)} = -5.1$; $p < .001$). Response bias is not a consideration in forced-choice because every recognition trial contains a target item (Macmillan & Creelman, 1991).

Subgroup Analysis of False Alarms on Yes/no Recognition

The number of false alarms in MCI was variable across subjects, ranging from 0 to 14 (Figure 2B). MCI subjects could be evenly separated into subgroups having relatively

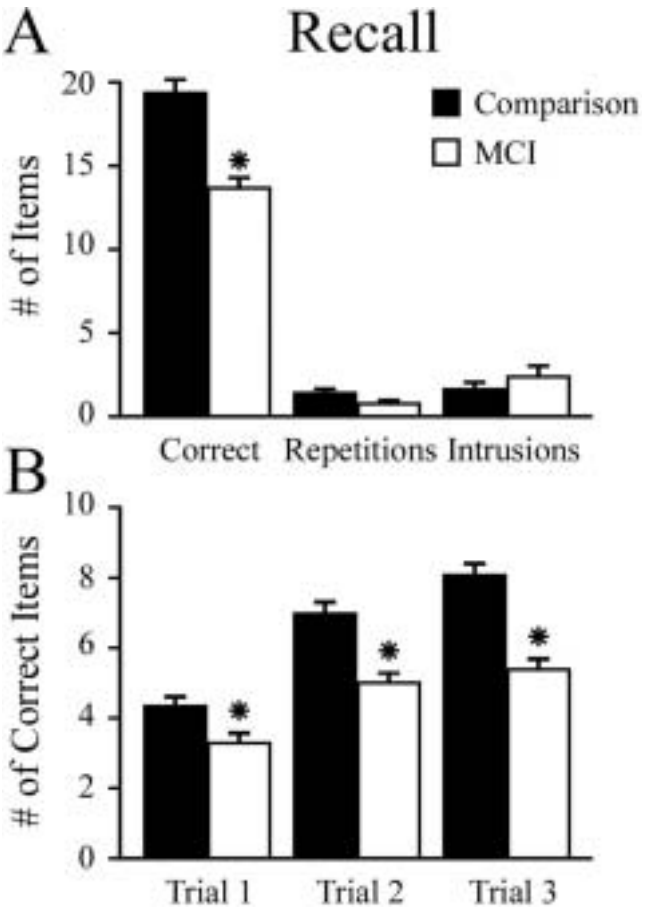

Figure 1. (A) The total number of items correctly recalled, repetitions, and intrusions for MCI and comparison subjects. MCI subjects recalled significantly fewer items than comparison subjects, with no group differences in repetitions or intrusions. (B) The mean number of items correctly recalled for each recall trials is shown for comparison and MCI groups. For both groups, the number of items recalled increased across trials. Asterisk indicates a significant group difference ($p < .05$).

low (0–5, M = 2 ± 2) or high (6–14, M = 10 ± 3) numbers of false alarms. The number of false alarms in the low false alarm MCI subgroup was not significantly different from the comparison group. Comparisons between MCI subjects having low versus high numbers of false alarms indicated significant differences in hits ($t_{(19)} = 2.3$; $p < .04$), d' ($t_{(19)} = 4.2$; $p < .001$), and c ($t_{(19)} = 5.2$; $p < .001$) for yes/no recognition. D' on forced-choice was also significantly lower in the high false alarm subgroup ($t_{(19)} = 2.2$; $p < .05$). Compared to the lower false alarm subgroup, MCI subjects with many false alarms had fewer hits, smaller d' values, and adopted a more lenient response criterion, indicating a greater propensity to respond "yes". Free recall performance (number correct) was not significantly different for the low (13.4 ± 0.9) and high (13.9 ± 0.9) false alarms subgroups, and there were also no significant differences in age or education among the MCI subgroups.

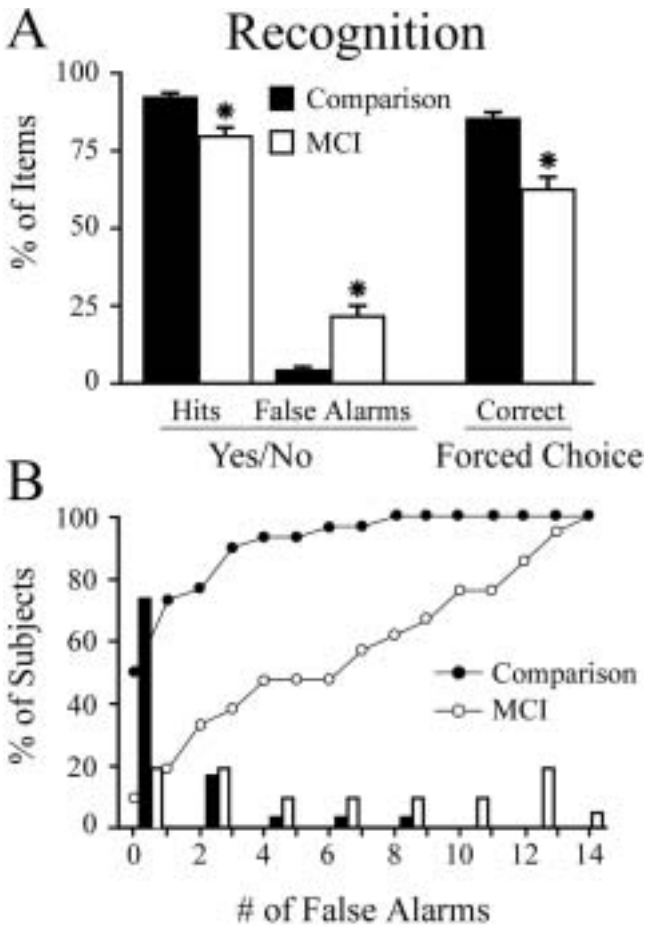

Figure 2. (A) Percent of yes/no recognition hits and false alarms, and correct items for forced-choice for MCI and comparison groups. MCI subjects recognized significantly fewer items in the yes/no (hits) and forced-choice (correct) tests, and produced more false alarms in the yes/no test than comparison subjects. Asterisk indicates significant group difference ($p < .05$). (B) Frequency and cumulative histograms of false alarms in comparison and MCI groups. Frequency histogram, illustrated by vertical bars, shows the percentage of subjects in each bin of two false alarm values (0–1, 2–3, etc.). Cumulative histograms are shown by the line graphs. Each point of the cumulative histogram shows the percentage of subjects with false alarm values that were less than or equal to the number of false alarms indicated on the x axis.

Discriminant Function Analysis

The ability of each test to classify subjects as either an MCI or comparison was assessed using stepwise discriminant analysis. Measures included Pr (proportion hits – proportion false alarms) from yes/no recognition, number correct from forced-choice recognition, and total number correct in free recall. Results showed that the best predictor of group membership was the Pr measure in the yes/no test, and inclusion of the other measures did not significantly improve the model (Wilks' $\lambda = 0.37$; $\lambda^2 = 47.6$; $p < 0.001$). The Pr measure correctly classified subjects with 93.3% sensitivity and 81% specificity. In contrast, total

number correct in free recall had 80% sensitivity and 66.7% specificity, and number correct in forced-choice had 86.7% sensitivity and 66.7% specificity.

Effect sizes using partial η^2 were calculated from separate univariate ANOVAs comparing groups for measures in yes/no recognition (Pr), forced-choice recognition (number correct), and free recall (total number correct). The effect size in yes/no ($\eta^2 = .63$) was nearly twice as large as the effect sizes using forced-choice ($\eta^2 = .37$) or free recall ($\eta^2 = .36$), indicating greater separation between groups in yes/no recognition. Taken together, the discriminant function and effect size results indicate that among the three tests that were compared yes/no recognition best distinguished MCI from comparison group.

Discussion

In this study free recall, yes/no recognition, and forced-choice recognition were examined in amnestic MCI and comparison groups. As expected, MCI subjects recalled fewer items than comparison subjects (Petersen et al., 1999; Smith et al., 1996). There were no significant group differences in repetitions, intrusions, recall consistency, subjective organization, and serial position effects. Relative to the comparison group, MCI subjects correctly recognized fewer study items during the yes/no and forced-choice tests, and produced more false alarms in the yes/no test. Recognition performance in the yes/no format classified individuals according to group (comparison, MCI) more accurately than either forced-choice or free recall measures. Group differences in recall and recognition memory performance and their implications for assessing memory functions in MCI are discussed below.

Recall Performance

MCI subjects recalled significantly fewer items that comparison subjects. This result was expected because other free recall tests were used to diagnose the episodic memory deficit in these MCI subjects, and is consistent with previous studies showing impaired recall in MCI (Collie & Maruff, 2000; Morris et al., 2001).

Recall performance exhibits regularities across multiple recall trials and between encoding and retrieval (Tulving, 1962). Items recalled on one trial are more likely to be recalled on subsequent trials (recall consistency; Bousfield & Bousfield, 1966); items recalled in sequential order are likely to be recalled together on subsequent trials (subjective organization; Tulving, 1962); and items memorized at the beginning or end of a list are recalled better than items in middle positions (serial position effects; Murdock, 1962). In this study there were no significant group differences on any of these three measures. Recall consistency is preserved during normal aging (Parker et al., 2004), although age-related declines were present when recall consistency was calculated using a different formula (Brown & Mitchell, 1991). Previous studies report that subjective organization declines in normal aging relative to young subjects (Davis et al., 2003; Witte, Freund & Sebby, 1990), and in subjects having frontal lobe lesions (Gershberg & Shimamura, 1995; Stuss, Craik Sayer, Franchi & Alexander, 1996). It is unknown if recall consistency or subjective organization are impaired in Alzheimer's disease. However, prefrontal neuropathology in early Alzheimer's disease (Morrison, Hof & Bouras, 1991; Price & Morris, 1999) suggests that, in contrast to MCI, deficits in subjective organization may be present in Alzheimer's disease. The absence of group differences in serial position effects is consistent with previous studies showing that serial position effects for short lists are preserved in early Alzheimer's disease (Bayley, et al., 2000; Burkart, Heun, Benkert, 1998).

Taken together, the present findings do not support the notion that free recall deficits in MCI are attributable to impaired organizational strategies.

Recognition Tests (Yes/no vs. Forced-choice)

MCI subjects were impaired in yes/no and forced-choice recognition tests relative to comparison subjects. For the yes/no test MCI subjects correctly recognized significantly fewer items and made more false alarms. Signal detection analysis suggested that overall group differences in hits and false alarms were not attributable to differences in response criteria. However, a subgroup of MCI subjects having many false alarms did have a more lenient response criterion, indicating a response bias in some MCI subjects (see below). Sensitivity measures for both yes/no and forced-choice tests indicate that MCI subjects are less able to distinguish targets from distractors, relative to the comparison group.

Discriminant function analysis showed that yes/no recognition was more accurate than forced-choice recognition in accurately classifying individuals according to group membership. Similarly, the effect size for the group difference was larger in yes/no versus forced-choice recognition. Greater deficits on yes/no versus forced-choice tests are also observed in amnesic patients with bilateral hippocampal and medial-temporal lobe lesions (Holdstock et al., 2002 cf. Khoe et al., 2000), which suggests that yes/no recognition may be more dependent on medial-temporal lobe function than forced-choice recognition. Early Alzheimer's disease pathology in medial temporal lobe structures, such as the entorhinal cortex (Gomez-Isla et al., 1996; Kordower et al., 2001; Price & Morris, 1999), may be associated with recognition deficits in MCI.

It is important to note that differences in the magnitude of the group effects for yes/no versus forced-choice recognition may be affected by differences in the number of study trials for each type of test. In yes/no recognition subjects were given three study-test trials of the free recall test, while in forced-choice recognition subjects memorized the study list once. Multiple presentations of the target words in the yes/no test would likely improve performance, compared to the single study trial used in forced-choice testing in both comparison and MCI groups. These results were not observed in the current study, however further research is required to determine the effect of different amounts of study trials on yes/no and forced-choice performance.

False Alarms

In the yes/no test MCI subjects incorrectly identified distractor items as targets (false alarms) more often than comparison subjects. The number of false alarms in the MCI group was variable across subjects. Eight MCI subjects had > 8 false alarms, seven performed similar to comparison subjects (0–2 false alarms), and the remainder had 3–7 false alarms (Figure 2B). Signal detection theory analysis showed that MCI and comparison groups used similar response criterion, indicating that increased false alarms were not due to MCI subjects having a response bias. However, when the MCI group was evenly divided into two subgroups, MCI subjects with relatively high false alarm rates (6–14 false alarms) had a more lenient response criterion. The lenient response criterion in the subgroup of MCI with many false alarms suggests a bias to respond "yes" compared to MCI subjects with fewer false alarms (0–5 false alarms). High false alarm rates are associated with frontal lobe damage because false alarms are commonly seen in patients with Alzheimer's disease (Gainotti, Marra, Villa, Parlato & Chiarotti, 1998) and frontal lobe lesions (Daum & Mayes, 2000). Additional study is needed to determine if MCI subjects

having a high number of false alarms have an even greater risk of developing early Alzheimer's disease, relative to MCI subjects with fewer false alarms.

Clinical Utility of Recognition Tests (USC-REMT)

In this study amnestic MCI subjects recalled significantly fewer items during free recall relative to comparison subjects. Recognition performance also showed significant group differences for yes/no recognition hits (MCI < comparison), yes/no recognition false alarms (MCI > comparison), and forced-choice recognition correct (MCI < comparison). In addition, discriminant function analysis revealed yes/no recognition was a better predictor of group membership, and had a larger effect size, compared to free recall or forced-choice tests.

Although recall tests are traditionally used to assess memory function in MCI (Collie & Maruff, 2000), recognition tests may also provide useful information for the purpose of detecting early Alzheimer's disease in MCI (Branconnier, Cole, Spera & DeVitt, 1982). Compared with recall, recognition performance declines less in normal aging (Parker et al., 2004; Zakzanis et al., 1999; Borke & Light, 1981), which can be an advantage when attempting to identify age-related neurological disorders such as MCI and Alzheimer's disease. Recall performance also declines in a variety of neurological disorders, such as vascular dementia, Parkinson's disease, and psychiatric conditions such as depression, which are common in older patients (Hodges, 2000; Zakzanis et al., 1999). Thus, recognition measures may help enhance diagnostic specificity in defining MCI and early Alzheimer's disease.

Recall and recognition measures can be sensitive and specific to episodic memory decline in Alzheimer's disease (Gainotti et al., 1998). Previous studies report that cued recall and paired associate learning correctly classified Alzheimer's disease versus healthy controls (Buschke, Sliwinski, Kuslansky & Lipon, 1997; Lindebcom, Schmord, Tulner, Washtra & Jonkes, 2002). Moreover, studies that manipulate encoding and/or retrieval conditions may increase sensitivity for detecting memory decline in early Alzheimer's disease using recall measures. Well-established memory tests such as the Consortium to Establish a Registry for Alzheimer's Disease (CERAD; Morris, Mohs, Rogers, Fillenbaum & Heyman, 1988), California Verbal Learning Test (CVLT; Delis, Kramer Kaplan & Ober, 1987), and Hopkins Verbal Learning Test (HVLT; Brandt, 1991) include a recognition subtest. The advantage of a screening test such as the USC-REMT is that it is a brief (~10 minute) memory assessment with two different recognition tests that showed minimal ceiling effects in healthy older controls. The USC-REMT also has seven memory lists that are equivalent for recall and recognition tests, which can be used over multiple follow-up visits. Furthermore, the yes/no test alone classified MCI subjects with 93% sensitivity 81% and specificity, which is comparable to classification of MCI patients using recall scores (Barbeau et al., 2004; Loewestein et al., 2004).

The current study was designed to assess memory performance in small set of individuals with amnestic MCI relative to an older comparison group. Results indicate that in the USC-REMT, yes/no recognition may be better at identifying memory impairment in MCI and early Alzheimer's disease compared to free recall and forced-choice recognition. Because the current study was conducted with a limited sample, future studies are necessary to replicate these findings with a larger group of subjects. Additionally, comparing the effect in an Alzheimer's disease group and in other subtypes of MCI would provide more information about group differences across the tests.

References

APA. (1994) *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association.

Backman, L., Small, B., Fratiglioni, L. (2001) Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain, 124*, 96–102.

Barbeau, E., Didic, M., Tramoni, E., Felician, O., Joubert, S., Sontheimer, A. et al. (2004) Evaluation of visual recognition memory in MCI patients. *Neurology, 62*, 1371–1322.

Bastin, C., Van der Linden, M. (2003) The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology, 17*(1), 14–24.

Bayley, P.J., Salmon, D.P., Bondi, M.W., Bui, B.K., Olichney, J., Delis, D.C. et al. (2000) Comparison of the serial position effect in very mild Alzheimer's disease, mild Alzheimer's disease, and amnesia associated with electroconvulsive therapy. *J Int Neuropsychol Soc, 6*, 290–298.

Bousfield, A.K., Bousfield, W.A. (1966) Measurement of clustering and of sequential constancies in repeated free recall. *Psychol Rep, 19*(3), 935–942.

Branconnier, R.J., Cole, J.O., Spera, K.F., DeVitt, D.R. (1982) Recall and recognition as diagnostic indices of malignant memory loss in senile dementia: A bayesian analysis. *Exp Aging Res, 8*(4), 189–193.

Brandt, J. (1991) The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *Clin Neuropsychol, 5*, 125–142.

Brown, A.S., Mitchell, D.B. (1991) Age differences in retrieval consistency and response dominance. *J Gerontol, 46*(6), 332–339.

Burkart, M., Heun, R., Benkert, O. (1998) Serial position effects in dementia of the Alzheimer type. *Dement Geriatr Cogn Disord*, 9, 130–136.

Burke, D.M., Light, L.L. (1981) Memory and aging: The role of retrieval processes. *Psychol Bull, 90*(3), 513–546.

Buschke, H., Sliwinski, M., Kuslansky, G., Lipton, R.B. (1997) Diagnosis of early dementia by the double memory test: Encoding specificity improves diagnostic sensitivity and specificity. *Neurology, 48*(4), 989–997.

Christensen, H., Kopelman, M.D., Stanhope, N., Lorentz, L., Owen, P. (1998) Rates of forgetting in Alzheimer's dementia. *Neuropsychologia*, 36(6), 547–557.

Collie, A., Maruff, P. (2000) The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment. *Neurosci Biobehav Rev 24*(3), 365–374.

Craik, F.I.M., McDowd, J.M. (1987) Age differences in recall and recognition. *J Exp Psychol Learn Mem Cog, 13*(3), 474–479.

Daum, I., Mayes, A.R. (2000) Memory and executive function impairments after frontal or posterior cortex lesions. *Behav Neurol, 12*, 161–173.

Davis, H.P., Small, S.A., Stern, Y., Mayeux, R., Feldstein, S.N., Keller, F.R. (2003) Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex*, 39, 1063–1091.

Delis, D.C., Kramer, J.H., Kaplan, E., Ober, B.A. (1987) *The California Verbal Learning Test*. New York: The Psychological Corporation.

Flicker, C., Ferris, S.H., Reisberg, B. (1991) Mild cognitive impairment in the elderly: Predictors of dementia. *Neurology 41*(7), 1006–1009.

Folstein, M.F., Folstein, S.E., McHugh, P.R. (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res, 12*(3), 189–198.

Gainotti, G., Marra, C., Villa, G., Parlato, V., Chiarotti, F. (1998) Sensitivity and specificity of some neuropsychological markers of Alzheimer dementia. *Alzheimer Dis Assoc Disord, 12*(3), 152–162.

Galasko, D., Bennett, D., Sano, M., Ernesto, C., Thomas, R., Grundman, M., et al. (1997) An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord, 11(Suppl 2)*, S33–S39.

Gershberg, F.B., Shimamura, A.P. (1995) Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia, 13*(10), 1305–1333.

Golob, E.J., Johnson, J.K., Starr, A. (2002) Auditory event-related potentials during target detection are abnormal in mild cognitive impairment. *Clin Neurophysiol, 113*(1), 151–161.

Gomez-Isla, T., Price, J.L., McKeel, D.W., Jr., Morris, J.C., Growdon, J.H., Hyman, B.T. (1996) Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *J Neurosci, 16*(14), 4491–4500.

Green, D.M., Swets, J.A. (1966) *Signal detection theory and psychophysics*. New York: John Wiley and Sons, Inc.

Hodges, J.R. (2000) Memory in the dementias. In E. Tulving & F.I., Craik (ed), *The Oxford handbook of memory*. New York: Oxford University Press, Inc.; 441–459.

Hodges, J.R., Patterson, K. (1995) Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia, 33*(4), 441–459.

Holdstock, J.S., Mayes, A.R., Roberts, N., Cezayirli, E., Isaac, C.L., O'Reilly, R.C. et al. (2002) Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus*, 12, 341–351.

Kaplan, E., Snodgrass, H., Weintraub, S. (1983) *Boston Naming Test*. Philadelphia, PA: Lea and Febiger.

Khoe, W., Kroll, N.E.A., Yonelinas, A.P., Dobbins, I.G., Knight, R.T. (2000) The contribution of recollection and familiarity to yes-no and forced-choice recognition tests in healthy aging subjects and amnesics. *Neuropsychologia*, 38, 1333–1341.

Kordower, J.H., Chu, Y., Stebbins, G.T., Dekosky, S.T., Cochran, E.J., Bennett, D.A. et al. (2001) Loss and atrophy of layer II entorhinal cortex neurons in elderly people with mild cognitive impairment. *Ann Neurol, 49*(2), 202–213.

Kroll, N.E.A., Yonelinas, A.P., Dobbins, I.G., Frederick, C.M. (2002) Separating sensitivity from response bias: implications of comparisons of yes-no and forced-choice tests for models and measures of recognition memory. *J Exp Psychol Gen, 131*(2), 241–254.

Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., Jonker, C. (2002) Visual association test to detect early dementia of the Alzheimer's type. *J Neurol Neurosurg Psychiatry, 73*, 126–133.

Loewenstein, D.A., Acevedo, A., Luis, C., Crum, T., Barker, W.W., Duara, R. (2004) Semantic interference deficits and the detection of mild Alzheimer's disease and mild cognitive impairment without dementia. *J Int Neuropsychol Soc, 10*, 91–100.

Machulda, M.M., Ward, H.A., Borowski, B., Gunter, J.L., Cha, R.H., O'Brien, P.C. (2003) et al. Comparison of memory fMRI response among normal, MCI, and Alzheimer's patients. *Neurology*, 61, 500–506.

Macmillan, N.A., Creelman, C.D. (1991) *Detection theory: A user's guide*. New York: Cambridge University Press.

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.

Morris, J.C., Mohs, R.C., Rogers, H., Fillenbaum, G., Heyman, A. (1988) Consortium to establish a registry for Alzheimer's disease (CERAD) clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacol Bull, 24*(4), 641–652.

Morris, J.C., Storandt, M., Miller, P., McKeel, D.W., Jr., Price, J.L., Rubin, E.H. (2001) et al. Mild cognitive impairment represents early-stage Alzheimer's disease. *Arch Neurol, 58*, 397–405.

Morrison, J.H., Hof, P.R., Bouras, C. (1991) An anatomic substrate for visual disconnection in Alzheimer's disease. *Ann N Y Acad Sci*, 640, 36–43.

Moscovitch, M., Winocur, G. (1990) The neuropsychology of memory and aging. In: F.I. Craik, T.A. Salthouse (eds), *The handbook of aging and cognition*. Hillsdale: Lawrence Erlbaum Associates, 315–372.

Murdock, B.B. (1962) The serial position effect of free recall. *J Exp Psychol Gen, 64*(5), 482–488.

Parker, E.S., Eaton, E.M., Whipple, S.C., Heseltine, P.N., Brigde, T.P. (1995) University of Southern California repeatable episodic memory test. *J Clin Exp Neuropsychol, 17*(6), 926–936.

Parker, E.S., Landau, S.M., Whipple, S.C., Schwartz, B.L. (2004) Aging, recall, and recognition: A study on the sensitivity of the University of Southern California repeatable episodic memory test (USC-REMT). *J Clin Exp Neuropsychol, 26*(3), 428–440.

Petersen, R.C. (2003) *Mild cognitive impairment: Aging to Alzheimer's disease*. New York: Oxford Press.

Petersen, R.C. (2004) Mild cognitive impairment as a diagnostic entity. *J Intern Med, 256*, 183–194.

Petersen, R.C., Doody, R.S., Kurz, A., Mohs, R., Morris, J.C., Rabins, M.D. et al. (2001) Current concepts in mild cognitive impairment. *Arch Neurol*, 58, 1985–1992.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E. (1999) Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol 56*(3), 303–308.

Price, J.L., Morris, J.C. (1999) Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Ann Neurol 45*(3), 358–368.

Reitan, R.M. (1958) Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills* 8, 271–276.

Smith, G.E., Petersen, R.C., Parisi, J.E., Ivnik, R.J., Kokmen, E., Tangalos, E.G. (1996) et al. Definition, course, and outcome of mild cognitive impairment. *Aging Neuropsychol Cogn, 3*(2), 141–147.

Spreen, O., Benton, A.L. (1977) *Neurosensory Center Comprehensive Examination for Aphasia*. Victoria, B.C.: Neuropsychology Laboratory, University of Victoria.

Sternberg, R.J., Tulving, E. (1977) The measurement of subjective organization in free recall. *Psychol Bull, 84*(3), 539–556.

Stuss, D.T., Craik, F.I.M., Sayer, L., Franchi, D., Alexander, M.P. (1996) Comparison of older people and patients with frontal lesions: Evidence from word list learning. *Psychol Aging, 11*(3), 387–395.

Tabachnick, B.G., Fidell, L.S. (1996) *Using multivariate statistics*. 3rd ed. New York: Harper and Row.

Tulving, E. (1962) Subjective organization in free recall of "unrelated" words. *Psychol Rev, 69*(4), 344–354.

Wang, Q., Zhou, J. (2002) Retrieval and encoding of episodic memory in normal aging and patients with mild cognitive impairment. *Brain Res, 924*, 113–115.

Wechsler, D. (1997) *Wechsler adult intelligence scale*. 3rd ed. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997) *Wechsler memory scale*. 3 ed. San Antonio, TX: The Psychological Corporation.

Witte, K.L., Freund, J.S., Sebby, R.A. (1990) Age differences in free recall and subjective organization. *Psychol Aging*, 5(2), 307–309.

Zakzanis, K.K., Leach, L., Kaplan, E. (1999) *Neuropsychological differential diagnosis*. Lisse: Swets & Zitlinger.