ELSEVIER

Alzheimer's & Dementia ■ (2016) 1-9

Alzheimer's & Dementia

Featured Article

# The validity of the Memory Alteration Test and the Test Your Memory test for community-based identification of amnestic mild cognitive impairment

Seline Ozer[a,b,*], Krist Noonan[c], Melanie Burke[b], John Young[a], Sally Barber[a], Anne Forster[a], Roy Jones[c]

[a]*Academic Unit of Elderly Care and Rehabilitation, University of Leeds, Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, United Kingdom*

[b]*School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom*

[c]*Research Institute for the Care of Older People (RICE), Bath, United Kingdom*

**Abstract**

**Introduction:** This study investigated the validity of two brief cognitive tests (Memory Alteration Test [M@T] and Test Your Memory [TYM] test) for identifying people with aMCI in the community.
**Methods:** Older people were invited to participate by their general practitioner practice. Eligible participants were assessed for aMCI using an operationalized approach to the Petersen criteria and the M@T and TYM.
**Results:** Both tests demonstrated significant ability in discriminating between people with aMCI and controls (AUC = 0.91 for M@T and 0.80 for TYM [$P < .001$ for both]). M@T performed with higher sensitivity than TYM (85% vs. 63%) and similar specificity (84% vs. 87%). Both tests demonstrated moderate test-retest reliability ($\kappa = \sim 0.5$) and took <10 minutes to administer.
**Discussion:** M@T and TYM are quick to administer. M@T demonstrated higher diagnostic test accuracy than TYM and could provide an efficient method for identifying aMCI in clinical and research settings.


*Keywords:* Dementia; Alzheimer's disease; Mild cognitive impairment; Validity; Reliability; Neuropsychological assessment; Diagnostic test accuracy

## 1. Introduction

There has been a growing clinical and research interest in the early identification of people at risk of developing dementia. Mild cognitive impairment (MCI) has emerged as a term to capture the predementia phase of cognitive dysfunction [1] and is defined as "cognitive decline greater than that expected for an individual's age and education level but that does not interfere notably with activities of daily life" [2]. The amnestic form of MCI (aMCI), where the predominant symptom is memory impairment, is associated with elevated rates of conversion to Alzheimer's disease (AD) [3]. It has been suggested that it may be more effective to target interventions at people in this predementia phase of AD, before the progressive disease is established [4].

Amnestic MCI is however largely unrecognized in primary care as its diagnosis depends on complex neuropsychological assessment methods not usually available in this setting. There is a need for simple, quick, and sensitive cognitive tests that will provide a more efficient way of identifying people with aMCI. These would provide a useful resource to busy primary health care staff who are encouraged, as stated in UK national guidance, to refer people

The author declared no conflicts of interest.

*Corresponding author. Tel.: +44 1274 383908; Fax: +44 1274 382766.

E-mail address: seline.ozer@bthft.nhs.uk

http://dx.doi.org/10.1016/j.jalz.2016.03.014

who show signs of MCI for further assessment by memory assessment services to aid early identification of dementia [5]. They could also be applied by researchers to find suitable participants for enrollment into studies of candidate interventions targeted at this early stage of cognitive decline.

A recent systematic review found that over 40 brief cognitive tests have been developed and tested to identify people with aMCI [6]. Several of these cognitive tests demonstrated promising diagnostic test accuracy, although most studies were found to be at a high risk of bias due to the method of participant selection used. Most studies selected patients with known aMCI from memory clinics and compared their performance on the test under evaluation with an opportunistically recruited group of people assumed to have no cognitive impairment. This exposed the studies to risk of unblinding of the patient assessment process and potentially exaggerated diagnostic accuracy [7,8]. The present study aimed to address this limitation by assessing the validity of two brief cognitive tests in a cohort of participants all recruited from the community, without prior knowledge of their cognitive status, thereby reducing the risk of bias in the assessment process.

The Memory Alteration Test (M@T) and the Test Your Memory (TYM) test were selected for investigation in this study. The developers of M@T, which is a brief, interviewer-administered memory task, reported it to have very high sensitivity (96%) and high specificity (70%–79%) for discriminating between people with aMCI and healthy controls [9,10]. The developers of TYM reported that it had very high sensitivity (93%) and high specificity (86%) for discriminating between people with and without mild Alzheimer's disease [11]. A subsequent study using a Japanese version of the test highlighted its potential for use as a screening tool for aMCI, reporting high sensitivity (76%) and specificity (74%) [12]. The TYM has the added advantage of being self-administered and requiring minimal supervision.

The aim of the present study was to evaluate the effectiveness of the M@T and TYM for identifying people with aMCI by investigating: (1) their sensitivity and specificity in detecting aMCI in a community-based population in comparison with the widely used standard for diagnosing aMCI based on the Petersen criteria [13]; (2) their test-retest reliability performance; and, (3) their clinical utility, assessed in terms of administration time and completion rates.

## 2. Methods

### 2.1. Recruitment

Participants were recruited from nine Bradford, UK, general practitioner practices (total registered patient population of 85,870). Their primary care health records were first screened to identify people who (1) were aged 70 years and older; (2) were not resident in a care or nursing home; (3) did not have dementia; (4) did not have current depression; (5) did not have history of stroke within the previous 3 months; (6) were not receiving palliative care. Study information flyers were posted to these identified people. The flyer asked further eligibility questions and those who responded and met the following criteria (or required further clarification) were contacted by telephone: (1) self-reported difficulty with their memory; (2) spoke English; (3) had attended school for at least eight years; (4) had an informant available to answer some of the study questions. Additionally, 100 people who did not self-report memory difficulties were invited to take part. Further checks were carried out by phone to ensure that the person was medically stable and could travel to our research offices. The eligible volunteers subsequently gave informed written consent to participate and enrolled onto the study during a visit to their home. The study was approved by the Yorkshire and The Humber National Research Ethics Service Committee (ref: 12/YH/0207).

### 2.2. Assessment

We developed a standardized protocol of neuropsychological tests to objectively assess for cognitive impairment, and a classification consistent with aMCI was determined according to the Petersen criteria [13]. The tests encompassed the cognitive domains of: memory (California Verbal Learning Test (CVLT), 2nd Edition [14]), executive function and attention (Brixton Spatial Anticipation Test [15] & Trail Making Test Parts A & B [16]), visuospatial function (visual object and space perception, spatial subset [17], and Clock Drawing Test [16]) and language (Graded Naming Test [18] & Pyramids and Palm Trees test [19]). Activities of daily living (ADL) performance were assessed using the informant-administered Bristol Activities of Daily Living Scale [20]. In addition, the National Adult Reading Test [21] was administered to provide an indication of premorbid verbal IQ. Mood was also assessed, initially via two depression screening questions [22,23] included in the study information flyer. Later, these were removed from the flyer and a more detailed assessment of mood was completed using the Geriatric Depression Scale-short form (GDS) [24], once the participant had been enrolled onto the study. The GDS was administered to the majority (93%) of participants. As low mood is known to impact on memory performance, all those participants who scored $\geq$6 on the GDS were classified as having "low mood" and were excluded from further analyses.

Participants who demonstrated impairment in memory (defined as CVLT short delay and long delay free recall $\geq$ 1.5 standard deviations below mean of published norms) and no impairment in ADL were classified as aMCI. Both single-domain (memory impairment only) and multi-domain (memory impairment and one or more other cognitive domain impairment) aMCI participants were included.

The other possible classification categories after the neurocognitive assessment process were (1) nonamnestic MCI

(impairment in non-memory cognitive domain(s)) and (2) cognitive difficulties beyond MCI (impairment in one or more cognitive domains and impairment in ADL).

Participants who did not meet the criteria for aMCI, or the other possible study classifications, were classified as "controls" and formed the reference group for the discriminatory analyses.

### 2.3. Procedures

All participants were assessed using the M@T and the TYM administered in a randomized order to avoid order effects, and the standardized battery of neuropsychological tests described previously. One of the brief cognitive tests was administered during session 1 which took place in the participant's home. The other brief cognitive test followed by the neuropsychological test battery were administered during session 2 which took place in our research facility within two weeks of session 1. The M@T, TYM, and neuropsychological battery were all administered by research assistants who were blinded to each other's assessments. Both the M@T and TYM were timed by stopwatch. Classifications of participants were agreed in consensus with the study neuropsychologist (KN), who was blinded to the results of the M@T and the TYM. To assess test-retest reliability, session 3 was arranged for a sample of participants who were re-administered the brief cognitive test they had completed during session 1. Session 3 was scheduled to take place at home within four weeks of session 1.

### 2.4. Memory Alteration Test

The Memory Alteration Test (M@T) [10] is an interviewer-administered test comprising a minimum of 33, and a maximum of 43, questions depending on free recall success. It assesses five cognitive skills (encoding, orientation, semantic memory, free recall, and cued recall, with recall intervals of <10 minutes) with a maximum total score of 50. It was developed and validated in Spain but has been translated into English, although not validated in this form. The translated version from the development article was applied [10] (with slight amendments made to the wording of some of the semantic memory questions; see Supplementary Table 1).

### 2.5. Test Your Memory test

The Test Your Memory (TYM) test [11] is a supervised, self-completed questionnaire comprising ten cognitive tasks, providing assessment of a wider range of cognitive domains than is covered in the M@T. In addition to memory and orientation tasks, the TYM also includes calculation, fluency, similarities, naming, and visuospatial tasks. As with the M@T, the recall interval for the memory task is <10 minutes. A score of five is also given for the amount of help that the participant required to complete the task, with higher scores indicating that less support was required. The maximum total score is 50.

### 2.6. Statistical analysis

All analyses were performed using SPSS statistics v22 (IBM). Between-group differences (aMCI vs. control) in age, years of education, IQ, and the M@T and TYM scores were explored using the Mann–Whitney *U* test (because the data were non-normally distributed). Difference in gender proportion between the groups was analyzed using the $X^2$ test.

Receiver operating characteristic (ROC) curve analysis was applied to assess the ability of the M@T and TYM global and subtest scores to discriminate between the aMCI group and the control group for a range of cut-off values. The area under the curve (AUC) was reported as a single measure of overall accuracy. Optimal cut-off points were defined as those providing the highest Youden index (calculated as "sensitivity + (specificity − 1)") [25]. Positive and negative predictive values and likelihood ratios were calculated for each optimal cut-off point.

Test-retest reliability of the M@T and TYM was investigated using the established techniques of Bland and Altman [26]. The mean difference between original test and retest scores was calculated, as was the reliability coefficient, which is twice the standard deviation of the differences and provides a measure of random error. Paired sample *t* tests were applied to explore for any significant differences between original test and retest scores (because the data were normally distributed). Agreement between original test and retest classifications (based on optimal cut-off scores) was also explored using the kappa statistic (established categories for interpreting the kappa statistic were applied from poor (<0.00) to moderate (0.41–0.60) to almost perfect (0.81–1.00) [27].

## 3. Results

From 1477 initial responses to the flyer, 507 participants were recruited (see Fig. 1 for recruitment flow); 31 people withdrew after session 1; and four people could not be classified (two had visual/hearing impairment that affected their performance on the tasks and two could not complete the assessment due to distress or fatigue). Thus, 472 participants were assessed with the standardized battery of neuropsychological tests and classified. Seventy two percent of these participants completed session 2 within two weeks of session 1.

Of the 472 participants classified, 52 (11.0%) people had aMCI, 26 (5.5%) people had nonamnestic MCI, 14 (3.0%) people had cognitive difficulties beyond MCI, and 20 (4.2%) people had low mood. The remaining 360 (76.3%) people were designated controls and formed the reference group for the subsequent analyses. The average individual test scores from the standardized neuropsychological battery of tests for the aMCI and control groups are provided as Supplementary Material (see Supplementary Table 1).

The aMCI participants were significantly older, had fewer years of education and a lower IQ than the controls (see Table 1 and Supplementary Table 2). As these factors could

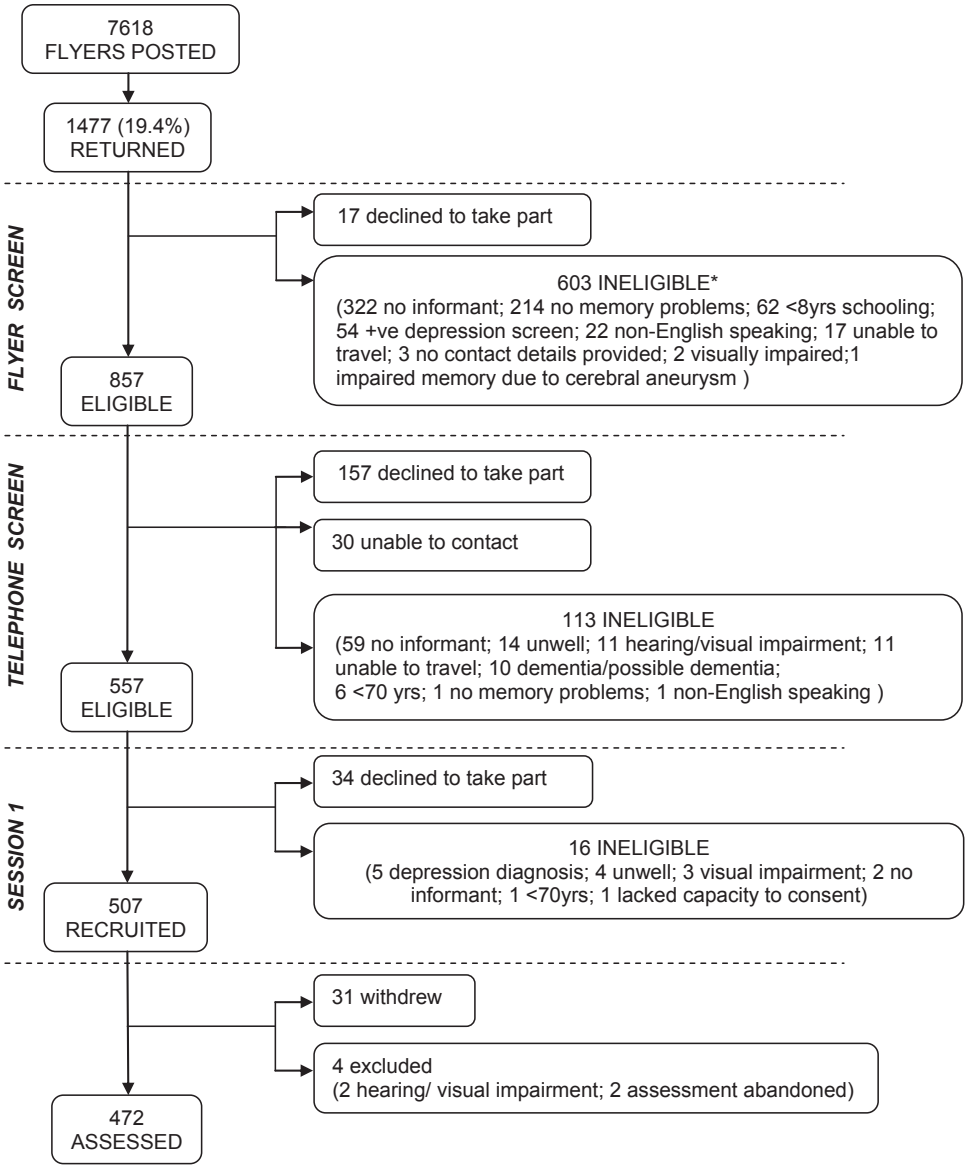


Fig. 1. Recruitment flow through the study. (*some respondents were ineligible for >1 reason).

Table 1
Demographic characteristics and M@T & TYM test scores for aMCI and Control participants

| Characteristic/Score | aMCI (n = 52) | Control (n = 360) | *P* value |
|---|---|---|---|
| Female, n (%) | 26 (50) | 162 (45) | .6* |
| Age, y | 78 (10) | 75 (7) | <.001 |
| Education, y | 11 (2) | 12 (4) | .001 |
| NART IQ | 112 (19) | 116 (13)[†] | .02 |
| M@T global score | 35 (10)[‡] | 45 (5)[§] | <.001 |
| TYM global score | 41 (8) | 47 (4)[†] | <.001 |

Abbreviations: NART, National Adult Reading Test; M@T, Memory Alteration Test; TYM, Test Your Memory test.

NOTE. Data are presented as median (interquartile range) unless indicated otherwise; *P* value for Mann–Whitney *U* test and *Chi-squared test.

[†]n = 359.

[‡]n = 51.

[§]n = 358.

have had an influence on the M@T and TYM scores, age-matched, education-matched, and IQ-matched controls were randomly selected for the discriminatory analyses. The aim was to select three matched controls for each aMCI participant; however, some aMCI participants had <3 matches (and some none at all), which resulted in 40 aMCI cases matched with 112 controls. The demographic characteristics of these matched participants are provided in Supplementary Table 3.

### 3.1. M@T performance

#### 3.1.1. Validity

Participants with aMCI scored significantly lower on the M@T than the control participants (35 (10) vs. 45 (5),

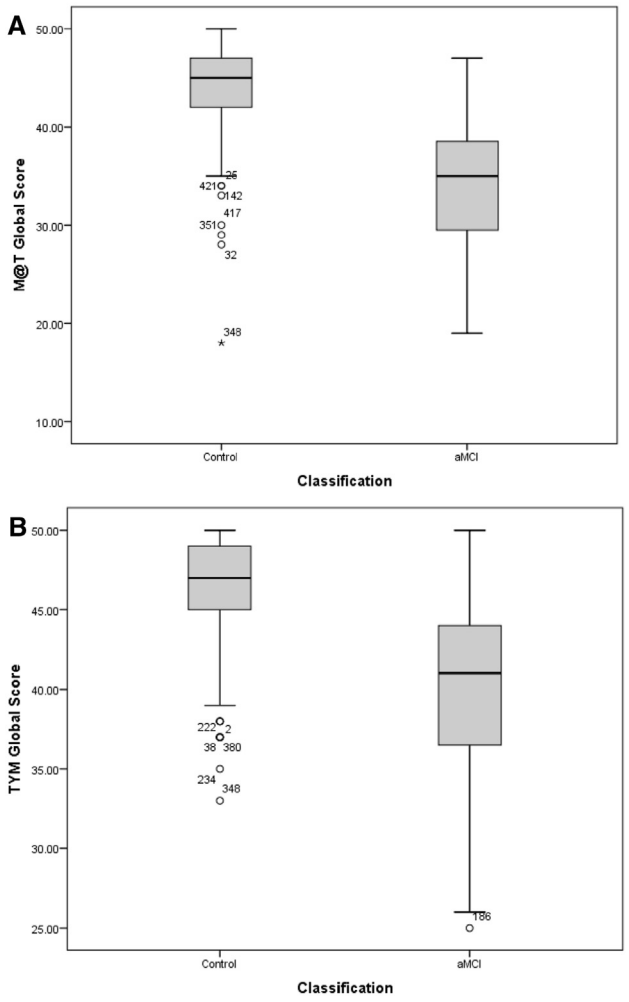


Fig. 2. (A) M@T and (B) TYM global scores for control and aMCI participants. Abbreviations: o, mild outliers (i.e., data points that lie >1.5 × IQR away from the lower/upper quartile); *, extreme outliers (i.e., data points that lie >3 × IQR away from the lower/upper quartile); IQR, interquartile range.

$U = 1459$, $z = -9.7$, $P < .001$, see Table 1). The box plots demonstrate the distribution of the M@T scores for each group (see Fig. 2A).

Fig. 3A shows the ROC curve of the M@T for differentiating the aMCI participants from the matched controls. The AUC was 0.91 and a score of 40 provided the optimal cut-off for discriminating between aMCI and controls (sensitivity 85%, specificity 84%; see Table 2). At the developer-recommended cut-off of 37 [10], a lower sensitivity (64%) but higher specificity (96%) was achieved.

The diagnostic utility parameters for the M@T subtests are also summarized in Table 2. The most sensitive subtests to discriminate between the aMCI and control groups were free recall and cued recall, which both demonstrated AUC >0.85. Orientation was the least sensitive subtest, with the lowest AUC value (0.61).

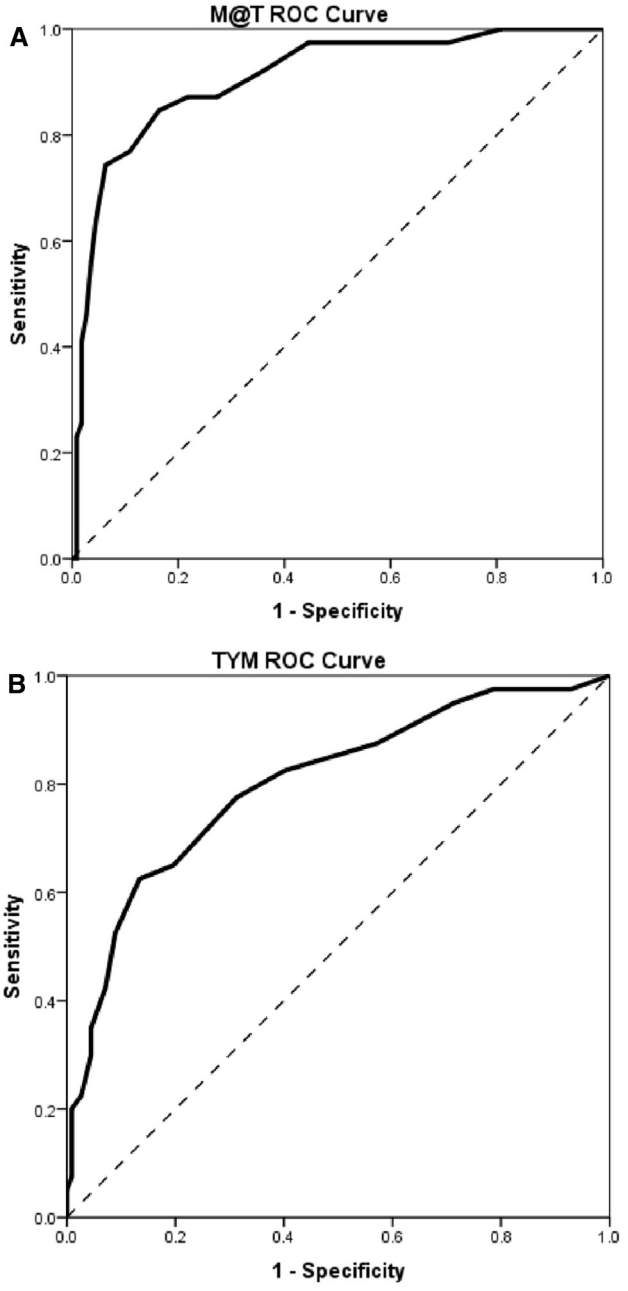


Fig. 3. Receiver operating characteristics of the (A) M@T and (B) TYM for differentiating aMCI participants from age, education, and IQ-matched controls.

### 3.1.2. Reliability

Twenty-five aMCI cases and 31 controls were reassessed with the M@T. Three quarters (75%) of these participants were reassessed within 4 weeks of session 1. Participants tended to score higher in session 3 than session 1 (mean difference, 2.8 points (95% CI = 2.0 to 3.7); see Table 3). This difference was significant (t (54) = −6.05, $P < .001$). The kappa value was 0.54 (indicating "moderate" agreement between sessions).

Table 2
Diagnostic utility of M@T and TYM to discriminate between aMCI and age, education and IQ-matched controls

| Test (maximum scores) | AUC (95% CI) | *P* value | Optimal Cut-off* | Sensitivity | Specificity | PPV | NPV | LR+ | LR− |
|---|---|---|---|---|---|---|---|---|---|
| M@T total score (50) | 0.91 (0.85–0.96) | <.001 | <40 | 85 | 84 | 37 | 98 | 5.31 | 0.18 |
| Encoding (10) | 0.79 (0.70–0.87) | <.001 | <9 | 58 | 88 | 35 | 95 | 4.83 | 0.48 |
| Orientation (5) | 0.61 (0.50–0.71) | .05 | <5 | 38 | 83 | 20 | 92 | 2.24 | 0.75 |
| Semantic (15) | 0.74 (0.65–0.83) | <.001 | <14 | 85 | 54 | 17 | 97 | 1.85 | 0.28 |
| Free recall (10) | 0.88 (0.82–0.94) | <.001 | <6 | 90 | 77 | 30 | 99 | 3.91 | 0.13 |
| Cued recall (10) | 0.86 (0.78–0.93) | <.001 | <9 | 80 | 78 | 29 | 97 | 3.64 | 0.26 |
| TYM Total score (50) | 0.80 (0.72–0.88) | <.001 | <43 | 63 | 87 | 35 | 95 | 4.85 | 0.43 |
| Orientation (10) | 0.57 (0.46–0.67) | .22 | <10 | 30 | 83 | 16 | 91 | 1.76 | 0.84 |
| Copying (2) | 0.52 (0.41–0.62) | .77 | <1 | 5 | 99 | 85 | 95 | 50.0 | 0.51 |
| Semantic (3) | 0.67 (0.56–0.77) | .002 | <2 | 40 | 91 | 33 | 93 | 4.44 | 0.66 |
| Calculation (4) | 0.58 (0.47–0.69) | .14 | <4 | 40 | 73 | 14 | 92 | 1.48 | 0.82 |
| Fluency (4) | 0.72 (0.63–0.82) | <.001 | <4 | 73 | 66 | 19 | 96 | 2.15 | 0.41 |
| Similarities (4) | 0.61 (0.51–0.72) | .04 | <4 | 53 | 68 | 16 | 93 | 1.66 | 0.69 |
| Naming (5) | 0.54 (0.43–0.65) | .46 | <5 | 13 | 96 | 27 | 91 | 3.25 | 0.91 |
| Visuospatial 1 (3) | 0.50 (0.40–0.61) | .99 | <1 | 13 | 94 | 19 | 91 | 2.17 | 0.93 |
| Visuospatial 2 (4) | 0.53 (0.42–0.63) | .65 | <4 | 15 | 90 | 14 | 91 | 1.50 | 0.94 |
| Free recall (6) | 0.72 (0.62–0.82) | <.001 | <3 | 50 | 93 | 44 | 94 | 7.14 | 0.54 |
| Help (5) | 0.53 (0.43–0.64) | .55 | <4 | 13 | 93 | 17 | 91 | 1.86 | 0.94 |

Abbreviations: PPV, positive predictive value; NPV, negative predictive values; LR+, positive likelihood ratio; LR−, negative likelihood ratio.
PPV and NPV calculated for 10% prevalence of MCI.
*Cut-off providing highest Youden index.

#### 3.1.3. Utility

The median time to complete the M@T in the control group was 6 min 5sec (IQR = 1 min 40 sec). Participants with aMCI took significantly longer than the control group, with a median time of 8 min 15 sec (±1 min 45 sec; U = 2798, z = −7.95, $P < .001$).

The majority (n = 409, 99%) of aMCI and control participants completed all the M@T questions. One aMCI participant had one missing item from the free recall subset, and two control participants each had one missing item from the cued recall subset. In addition, one aMCI participant had missing data from their retest M@T (three missing items from the cued recall subset).

The M@T requires the participant to encode and recall five words: cherry, axe, elephant, piano, and green. However, it was noted that the words "axe" and "green" were commonly misheard and had to be repeated.

### 3.2. TYM performance

#### 3.2.1. Validity

Participants with aMCI scored significantly lower on the TYM than the control participants (41 (8) vs. 47 (4), U = 2921.5, z = −8.1, $P < .001$, see Table 1). The box plots demonstrate the distribution of the TYM scores for each group (see Fig. 2B).

Fig. 3B shows the ROC curve of the TYM for differentiating the aMCI participants from the matched controls. The AUC was 0.80, and a score of 43 provided the optimal cut-off for discriminating between aMCI and controls (sensitivity 63%; specificity 87%; see Table 2). At the commonly used cut-off of 44 [12,28,29], a slightly higher sensitivity (65%) but lower specificity (80%) was achieved.

The diagnostic utility parameters for the TYM subtests are summarized in Table 2. All subtests (except for the Fluency subtest) performed with less sensitivity than the global TYM test, and all subtests demonstrated AUC values of <0.75. Fluency and free recall were the most accurate subtests, with AUCs of 0.72.

Table 3
Test-retest reliability of the M@T and TYM

| | Bland and Altman results | | | | |
|---|---|---|---|---|---|
| Test | Mean difference | 95% CI for mean difference | $SD_{diff}$ | Reliability coefficient | Cohen's kappa |
| M@T | −2.8 | −2.0 to −3.7 | 3.5 | 6.9 (of 50) | 0.54* |
| TYM | −1.9 | −1.0 to −2.8 | 3.0 | 6.0 (of 50) | 0.51* |

*$P < .001$.

#### 3.2.2. Reliability

Nineteen aMCI cases and 30 controls were reassessed with the TYM. The majority (88%) of these participants were reassessed within 4 weeks of session 1. Participants tended to score higher in session 3 than session 1 (mean difference = 1.9 points [95% CI = 1.0 to 2.8]; see Table 3). This difference was significant (t(48) = −4.40, $P < .0005$). The kappa value was 0.51 (indicating "moderate" agreement between sessions).

#### 3.2.3. Utility

The median time to complete the TYM in the control group was 7 min 19 sec (±2 min 30 sec). Participants with aMCI took significantly longer ($P < .005$) than the control group, with a median time of 9 min 26 sec (±2 min 32sec; U = 4374.5, z = −5.59, $P < .001$).

Fully completed TYM questionnaires were obtained for all the aMCI participants and almost all the controls. Only one control participant had missing items with four missing items from the orientation subtest and no score for the help given subtest.

## 4. Discussion

This study aimed to assess the accuracy of two brief cognitive tests (M@T and TYM) for identifying people with aMCI in the community. The M@T performed with higher diagnostic test accuracy than the TYM, with higher sensitivity (85% vs. 63%), similar specificity (84% vs. 87%), and higher overall accuracy as demonstrated by the AUC values (0.91 vs. 0.80). Both tests were associated with a learning effect such that a second assessment repeated within 1 month of the first-showed higher test scores. Both tests were acceptable to participants with completion times of <10 minutes and very few missing items.

Although the M@T demonstrated reasonably high levels of sensitivity and specificity for aMCI, the study did not reproduce the very high diagnostic test accuracy (DTA) results reported in previous studies. For example, a recent study by Custodio et al reported that a cut-off score of 37 had a sensitivity and specificity of 98% (AUC = 0.999) to differentiate aMCI from controls [30]. The developers of the M@T recommend a cut-off score of 37, and they report sensitivity of 96% and specificity of 70%–79% at this cut-off [9,10]. However, we found a higher optimal cut-off value for our sample (<40), and our data demonstrated a lower sensitivity (63%) but higher specificity (96%) at the recommended cut-off.

Our DTA results for TYM also differed from those reported in previous studies. A cut-off score of 44 has been recommended in three previous studies of TYM [12,28,29]. These studies report sensitivities of 74%–86% and specificities of 60%–74% at this cut-off. We found a slightly lower optimal cut-off value in our sample (<43) and lower sensitivity (65%) but higher specificity (80%) at the previously recommended cut-off.

These DTA discrepancies might be explained by our community-based recruitment method. The previous studies were all conducted in secondary/specialist care settings such as memory clinics [9,10,12], neurology departments [28,30], or psychiatry units [29] and most recruited their aMCI sample from patients attending clinics and their "control" sample from a separate source, such as other hospital departments or the wider community [10,12,29,30]. Studies which use a "case-control" design such as this are known to exaggerate diagnostic accuracy [7,8]. The present study used a sampling method that was designed to reflect how the brief cognitive tests might be applied in routine care in the future, that is, community-based aMCI case finding (refined by self-reported questions on memory difficulties). This approach has resulted in more conservative estimates of DTA which are likely to be more generalizable to unselected populations. Assessing all participants with the same reference standard also meant that we avoided verification bias which occurs when only a proportion of the study population receives confirmation of the diagnosis (usually those with positive test results) and can also result in overestimation of DTA values [7,8]. The fact that both tests were found to perform at lower sensitivity at the recommended cut-offs than previously demonstrated indicates that our aMCI population was less impaired than those included in previous studies, likely to be a result of our community-based, rather than secondary care-based, approach to recruitment.

As reported by the developers of M@T, our study also demonstrated that free recall and cued recall were the most accurate subtests for discriminating between the aMCI and control groups. It is perhaps unsurprising that these recall scores are the most useful for identifying aMCI because it is well known that episodic memory is impaired in aMCI and early AD [31]. This is thought to be the result of early pathologic changes that occur in the hippocampus and medial temporal lobe [32]. Similarly, the Free Recall subtest was found to be the most accurate TYM subtest for identifying people with aMCI. Some TYM subtests were found to be of less value in discriminating between aMCI and controls (e.g., the naming subtest and two visuospatial subtests, all with AUC values ~0.5). These findings indicate that the recall subtest scores are particularly useful for identifying aMCI and that particular emphasis should be placed on these subtests when interpreting scores.

It is important to note here that we applied an algorithmic, rather than clinical, categorization of aMCI. Although this differs from usual clinical practice, which would involve the incorporation of clinical and neurological examination to make a final diagnosis, it enabled us to apply the criteria in a standardized and objective manner, thereby ensuring reliability of our classifications [4]. The M@T and TYM have both demonstrated that they are valid in identifying people with aMCI as classified using this algorithmic method. Of course, in practice, further clinical assessment would be required to make a differential diagnosis, and it is the intention that these instruments would be used as a first screening stage in clinical practice and not as diagnostic tools.

Both the M@T and TYM performed with moderate test-retest reliability. Participants tended to score higher in the second session than the first on both tests indicating that there may have been a learning/practice effect. This is commonly seen with repeated cognitive testing [33]. The reliability coefficient (which reflects random error) was fairly high for both tests at 6.9 and 6.0 points for M@T and TYM, respectively. These values give an indication of the maximum change in score on retest that might be expected by chance in the absence of change in an individual's cognitive status. In other words, only a change of score that is more than seven points for M@T and six points for TYM would represent real change for an individual patient. To the

authors' knowledge, this is the first study to provide these data. This is relevant if these tests are to be used in applications such as measuring the effectiveness of interventions or monitoring change in cognition over time (although testing intervals may be longer than four weeks in these instances, which may lessen any practice effects).

Both tests were quick to administer, taking <10 minutes, with the M@T being slightly quicker than the TYM (by approximately 1 minute, on average). Furthermore, we found very little missing data for both tests indicating that there were no issues with administering them. Both tests were designed to be administered by nonspecialist staff (with the TYM requiring minimal supervision). A particular issue with the M@T arose concerning the words used to assess episodic memory. The words "axe" and "green" were often misheard by the participants and had to be repeated and so these may need to be replaced by more easily distinguishable words for use in English-speaking populations (e.g. "hammer", "yellow"). Any adaptations of the M@T would ideally need to be revalidated in further DTA studies.

A limitation of the present study is that there was no long-term follow-up of the participants and so the prognostic abilities of the tests cannot be commented on here. Future studies are required to see how accurate the tests are at discriminating between those people who go on to develop dementia and those who remain stable or improve. It would also be interesting to evaluate how the M@T and TYM might perform relative to other commonly used brief cognitive tests, such as the Memory Impairment Screen [34] and the Montreal Cognitive Assessment (MoCA) [35] in a similar setting, and head-to-head comparative studies are warranted in the future.

In summary, the present study has provided evaluation of the performance of M@T and TYM within a community-based UK setting, providing results that are generalizable to the wider population. Amnestic MCI is largely unrecognized in primary care due to the lack of simple, quick, and sensitive cognitive tests. Both M@T and TYM were simple and quick to use and demonstrated moderate test-retest reliability. However, M@T was found to perform with higher DTA than TYM and could provide an efficient and accurate method for identifying aMCI in clinical or research settings.

## Acknowledgments

The authors would like to thank the participants who gave their time to take part in the study. They also thank Dr Jeremy Brown for making the TYM test available for the study, and Dr Jose Luis Molinuevo and Dr Lorena Rami for giving permission to use the M@T and for providing advice on its administration. In addition, they would like to give thanks to Dr Jane Kirby for assisting with the recruitment of GP practices. Finally, they would like to thank the team of researchers who were involved with recruiting participants to the study and administering the assessments (Gillian Procter, Sarah Smith, Lesley Brown, Rosemary Shannon, Alex Benham, Katie Grenfell, Nicola Cornwall, and Jaime Smith).

This article presents independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0211-24044). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

## Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2016.03.014.

## RESEARCH IN CONTEXT

1. Systematic review: Over 40 brief cognitive tests have been developed to identify people with aMCI. However, these have mainly been tested in secondary care settings with highly selected patient and control groups. The Memory Alteration Test (M@T) and Test Your Memory (TYM) test have both demonstrated promising diagnostic test accuracy within this context. This is the first study aimed at investigating their performance in a community-based setting.
2. Interpretation: Our community-based approach to recruitment has resulted in more conservative estimates of diagnostic test accuracy for the M@T and TYM than previously reported. Nevertheless, the M@T in particular has demonstrated significant discriminative abilities and could provide an efficient and accurate method for identifying aMCI in clinical or research settings.
3. Future directions: The word list used to assess episodic memory in the M@T could be improved for English-speaking populations with any adaptations to be revalidated in future studies. Further studies are also required to investigate the prognostic abilities of the tests.

## References

[1] Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, et al. Mild cognitive impairment: ten years later. Arch Neurol 2009;66:1447–55.

[2] Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. Mild cognitive impairment. Lancet 2006;367:1262–70.

[3] Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, et al. Current concepts in mild cognitive impairment. Arch Neurol 2001; 58:1985–92.
[4] Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med 2014;275:214–28.
[5] National Institute for Health and Clinical Excellence. Dementia: Supporting people with dementia and their carers in health and social care, Clinical Guideline 42. London: NICE; 2006.
[6] Ozer S, Young J, Champ C, Burke M. A systematic review of the diagnostic test accuracy of brief cognitive tests to detect amnestic mild cognitive impairment. Int J Geriatr Psychiatry 2016. Advance online publication. http://dx.doi.org/10.1002/gps.4444.
[7] Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061–6.
[8] Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004;140:189–202.
[9] Rami L, Bosch B, Sanchez-Valle R, Molinuevo JL. The memory alteration test (M@T) discriminates between subjective memory complaints, mild cognitive impairment and Alzheimer's disease. Arch Gerontol Geriatr 2010;50:171–4.
[10] Rami L, Molinuevo JL, Sanchez-Valle R, Bosch B, Villar A. Screening for amnestic mild cognitive impairment and early Alzheimer's disease with M@T (Memory Alteration Test) in the primary care population. Int J Geriatr Psychiatry 2007;22:294–304.
[11] Brown J, Pengas G, Dawson K, Brown LA, Clatworthy P. Self administered cognitive screening test (TYM) for detection of Alzheimer' s disease: cross sectional study. BMJ 2009;338:b2030.
[12] Hanyu H, Maezono M, Sakurai H, Kume K, Kanetaka H, Iwamoto T. Japanese version of the Test Your Memory as a screening test in a Japanese memory clinic. Psychiatry Res 2011;190:145–8.
[13] Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004;256:183–94.
[14] Delis DC, Kramer JH, Kaplan E, Ober BA. California Verbal Learning Test. 2nd ed. San Antonio, TX: Psychological Corporation; 2000.
[15] Burgess PW, Shallice T. The Hayling and Brixton Tests. Bury St Edmonds: Thames Valley Test Company; 1997.
[16] Reitan RM, Wolfson D. The Halstead–Reitan Neuropsychological Test Battery: Therapy and clinical interpretation. Tucson, AZ: Neuropsychological Press; 1985.
[17] Warrington EK, James M. The Visual Object and Space Perception Battery. London, UK: Pearson Assessment; 1991.
[18] Warrington EK. The Graded Naming Test: A restandardisation. Neuropsychol Rehabil 1997;7:143–6.
[19] Howard D, Patterson K. Pyramids and palm trees: a test of semantic access from pictures and words. Bury St Edmonds: Thames Valley Test Company; 1992.
[20] Bucks RS, Ashworth DL, Wilcock GK, Siegfried K. Assessment of Activities of Daily Living in Dementia: Development of the Bristol Activities of Daily Living Scale. Age Ageing 1996;25:113–20.
[21] Nelson HE, Willison J. The National Adult Reading Test (NART). Windsor, UK: NFER-Nelson; 1991.
[22] National Institute for Health and Clinical Excellence. Depression in adults: The treatment and management of depression in adults, Clinical Guideline 90. London: NICE; 2009.
[23] Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener. Med Care 2003;41:1284–92.
[24] Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS): Recent Evidence and Development of a Shorter Version. Clin Gerontol 1986; 5:165–73.
[25] Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5.
[26] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
[27] Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33:159–74.
[28] Munoz-Neira C, Henriquez Chaparro F, Delgado C, Brown J, Slachevsky A. Test Your Memory-Spanish version (TYM-S): a validation study of a self-administered cognitive screening test. Int J Geriatr Psychiatry 2014;29:730–40.
[29] Szczesniak D, Wojtynska R, Rymaszewska J. Test Your Memory (TYM) as a screening instrument in clinical practice—the Polish validation study. Aging Ment Health 2013;17:863–8.
[30] Custodio N, Lira D, Herrera-Perez E, Nunez Del Prado L, Parodi J, Guevara-Silva E, et al. The Memory Alteration Test discriminates between cognitively healthy status, mild cognitive impairment and Alzheimer's disease. Dement Geriatr Cogn Dis Extra 2014;4:314–21.
[31] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–8.
[32] Jack CR, Shiung MW, Weigand SD, O'Brien PC, Gunter JL, Boeve BF, et al. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. Neurology 2005;65:1227–31.
[33] Heilbronner RL, Sweet JJ, Attix DK, Krull KR, Henry GK, Hart RP. Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: the utility and challenges of repeat test administrations in clinical and forensic contexts. Clin Neuropsychol 2010;24:1267–78.
[34] Buschke H, Kuslansky G, Katz M, Stewart WF, Sliwinski MJ, Eckholdt HM, et al. Screening for dementia with the memory impairment screen. Neurology 1999;52:231–8.
[35] Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53:695–9.