Neuropsychological Characterization of Patients with the WMT Dementia Profile

Anne Rienstra1,2,†*, Michelle Klein Twennaar1,2,†, Ben Schmand1,2

1Department of Brain and Cognition, University of Amsterdam, Amsterdam, The Netherlands
2Department of Neurology, Academic Medical Center, Amsterdam, The Netherlands

*Corresponding author at: Department of Brain and Cognition, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands.
Tel.: +31-20-525-6887; fax: +31 20 639 0531.
E-mail address: a.rienstra@uva.nl (A. Rienstra).
Accepted 5 April 2013

Abstract

The “dementia profile” is used to reduce false positives on the Word Memory Test (WMT). Provided that this profile reflects genuine memory impairment, corresponding cognitive deficits should be found in neuropsychological testing. We examined whether a WMT dementia profile is a significant indicator of cognitive impairment and/or decline. In addition, we evaluated the classification accuracy for the clinical diagnosis of dementia. Elderly patients (n = 167) with cognitive complaints were given an extensive neuropsychological test battery, including the WMT. This was repeated 2 years later. The results demonstrate that patients with the dementia profile have a higher chance of showing real cognitive impairment at baseline, and even more so 2 years later. They showed a faster cognitive decline than patients who passed the WMT effort subtasks. Sensitivity of the profile was a moderate 60%. However, the positive predictive value was high, viz. 81% at baseline and 93% at follow-up.

Keywords: Symptom validity; Effort testing; Word Memory Test; Cognitive impairment; Classification accuracy

Introduction

The Word Memory Test (WMT; Green, 2003) is a computerized word list-learning task designed to detect non-credible performance during neuropsychological evaluation. It is one of the most widely used symptom validity tests (SVTs; Hartman, 2002). The WMT has multiple symptom validity indices, hereafter referred to as “easy subtests” or “effort subtests,” and several conventional tests of verbal memory, hereafter called “hard subtests” (see “Methods” section for details). An examiner, who fails one of the easy subtests, is considered to exert non-credible performance. The easy subtests have high sensitivity and specificity in discriminating between persons asked to make a good effort and those instructed to feign memory impairment (Brockhaus & Merten, 2004; Iverson, Green, & Gervais, 1999; Jing, Slick, Strauss, & Hultsch, 2002). Moreover, compensation-seeking individuals were shown to suppress their performance on this task (Green, Lees-Haley, & Allen, 2002). Also, the WMT was found to be sensitive in differentiating between carefully defined malingering and non-malingering groups (Greve, Ord, Curtis, Bianchini, & Brennan, 2008).

A serious limitation of all SVTs, including the WMT, is that severe cognitive impairment may significantly interfere with performance. Some studies on the specificity of the WMT reported no differences in mean scores on the easy subtests between neurological patients with normal versus impaired memory (Green, Allen, & Astner, 1996). Even amnesic subjects with bilateral hippocampal damage scored above the WMT cut-offs (Goodrich-Hunsaker & Hopkins, 2009). However, other studies demonstrated lower specificity in severely impaired patients (Gorissen, Sanz, & Schmand, 2005; Greve et al., 2008; Merten, Bossink, & Schmand, 2007). Concern over the specificity of the WMT has also been raised in a recent meta-analytic review...
(Sollman & Berry, 2011). If a patient is giving abnormal but valid responses due to impaired but true abilities, he may erroneously be identified as exerting non-credible performance by a SVT. This false-positive error may lead to misdiagnosis and inappropriate clinical decision-making.

The WMT manual offers guidelines to allow examiners to differentiate non-credible responses from genuine memory impairment in case a patient failed the easy subtests (Green, 2003). This distinction is made with the “dementia profile.” Patients who show a dementia profile perform below the cut-off on the easy subtests, but their extremely low scores on the hard subtests suggest that this is due to genuine cognitive dysfunction. A difference of at least 30 percentage points between the mean of the easy subtests and the mean of the hard subtests defines the dementia profile (Green, 2003). Such a large easy–hard difference is usually absent in people with non-credible performance, whereas it is often seen in patients with dementia.

Most studies that examined the validity of profiles of scores across subtests to reduce false positives have used a short version of the WMT, the Medical SVT (MSVT; Green, 2004). On the MSVT, people suffering from dementia produced an easy–hard difference of at least 20 points. By this criterion, a maximum of 5% false positives was found in dementia patients (Howe, Anderson, Kaufman, Sachs, & Loring, 2007; Howe & Loring, 2009). Singhal, Green, Ashaye, Shankar, and Gill (2009) even reported that the MSVT had 100% specificity in a dementia group if the dementia profile was taken into consideration. Similar conclusions were drawn by Henry, Merten, Wolf, and Harth (2010) using profile analysis on a non-verbal variant of the MSVT, the NV-MSVT, in a group of neurological patients ($n = 65$), including a dementia subgroup ($n = 21$). No false positives were found in the dementia group, and the specificity in the whole neurological group was 97.5%. Green, Montijo, and Brockhaus (2011) tested persons with possible or probable dementia on the WMT and the MSVT to find the rate of false positives using the dementia profile. Combining the groups, they found a specificity of 98.4%, which represents a false-positive rate of 1.6%.

In summary, several studies have looked at the specificity of the dementia profile. In general, the findings show high specificity using different neurological groups.

The above-mentioned studies looked at the performance profiles of individuals with a specific neurological diagnosis. However, it is also interesting to look at the other way around (i.e., to look at the cognitive performance and diagnoses of patients with or without a dementia profile). In most cases, the dementia profile will be due to quite severe memory impairment. Consequently, it may convey important information on genuine dementia and its prodromal stage. Indeed, Howe and Loring (2009) reported that the MSVT dementia profile has a high positive predictive value (PPV; 89.5%) for dementia. To our best knowledge, so far, there is only one study that has addressed the analysis of the dementia profile on the MSVT (Axelrod & Schutte, 2010). Their findings show that individuals passing the easy subtests of the MSVT perform significantly better than patients failing the easy subtasks (either due to genuine cognitive impairment or to non-credible performance) on most tasks of a battery of neuropsychological tests assessing among other domains memory and executive functioning.

The aim of the present study was to look at neuropsychological functioning in patients with the WMT dementia profile. Provided that the dementia profile reflects genuine memory impairment, corresponding cognitive deficits should be found in neuropsychological testing. Even more so, given the high PPV of the dementia profile (Howe & Loring, 2009), it may contribute to predicting dementia and cognitive decline with progression of time. We therefore also explored the sensitivity and specificity of the dementia profile for the clinical diagnosis of dementia at baseline and at follow-up. Note that the terms sensitivity and specificity in this context carry another meaning than the usual in the context of symptom validity assessment. In the present context, we use the terms to refer to predictive validity for the diagnosis of cognitive impairment and dementia; here, they do not refer to non-credible responding.

A group of elderly participants with cognitive complaints was given a neurological evaluation and an extensive neuropsychological test battery, including the WMT. This was repeated 2 years later. Participants were divided into three groups according to their symptom validity performance on the WMT. We expected people with the dementia profile to perform worse on neuropsychological tests than the group with normal effort scores on the WMT. We also anticipated that the dementia profile group would show cognitive decline after 2 years and that this decline would be greater than in the patients with normal effort scores. Additionally, we hypothesized that the group showing non-credible performance would not show cognitive decline after 2 years, but might even improve in test performance.

Method

Participants

This study was part of the longitudinal project “Improving the early Diagnosis of Alzheimer’s Disease and Other dementias” (IDADO). Between February 2007 and October 2009, patients ($n = 170$) were recruited from the neurological and geriatric outpatient clinics, day clinics, and memory clinics of six general and psychiatric hospitals in the Netherlands (Academic Medical Center Amsterdam, Medical Center Alkmaar, Slotervaart Hospital Amsterdam, GGZ-Noord-Holland-Noord, and Geriant
Noord-Kennemerland). Patients between 50 and 85 years of age were included if they presented with (a) complaints of a decline in cognitive or behavioral functioning (as expressed by the patient or a close relative) not normal for age and (b) essentially intact instrumental activities of daily living. Patients were referred to the study by dementia specialists (neurologists, geriatricians, psychiatrists) after an initial consultation consisting of patient history, collateral information from a relative, clinical examination including a dementia screening instrument (Mini Mental State Examination [MMSE]; Folstein, Folstein, & McHugh, 1975), laboratory analyses, MRI or CT scan if deemed necessary, and application of the clinical dementia criteria, including the Clinical Dementia Rating (CDR; Morris, 1993). For all participants, the differential diagnosis included a possible early stage of dementia at the time of referral. Exclusion criteria were dementia as established during initial consultation by the dementia specialist according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV; American Psychiatric Association, 1994), other brain disease or systemic disease sufficient to cause the mental complaints, current substance abuse or addiction, a medical condition or handicap that precluded neuropsychological evaluation, pre-existent mental retardation, contraindications for MRI scanning, and insufficient command of the Dutch language. Psychiatric (co-)morbidity on Axis I according to DSM-IV (American Psychiatric Association, 1994), like major depressive disorder, common anxiety disorder, or adjustment disorder, was not an exclusion criterion.

Procedure

At baseline, participants underwent a comprehensive neuropsychological test battery to evaluate several cognitive domains, including two tests of symptom validity, administered by a neuropsychologist. Tests were administered in a fixed order. In intervals of verbal memory tests, other non-verbal neuropsychological tests were administered (and vice versa) in order to avoid interference. Furthermore, a structured psychiatric assessment was administered and a high-resolution 3-T structural brain MRI scan was made. Assessments were carried out in the hospital or at the participant’s home, in two sessions of 2–3 h with suitable rest periods. MRI scans were made within a month of the neuropsychological examination.

After 2 years, a neurologist again saw all patients. Next, a neuropsychologist or a trained neuropsychology master’s student, supervised by the neuropsychologist, administered a shortened neuropsychological test battery. This was done in one session of approximately 2 h. If possible, an MRI scan was made. Results and comparisons with baseline test results were reported back to the referring physician. For patients who could not participate in follow-up, final diagnostic information was obtained from the referring physician. The local ethics committees of participating hospitals approved the study. Written informed consent was obtained from all patients after the nature of the study was fully explained.

Neuropsychological Assessment

Global Cognitive Functioning. The MMSE (Folstein et al., 1975) is used to screen for cognitive impairment. It includes 30 questions and problems concerning orientation, memory, attention, verbal comprehension, and visuoconstructive abilities. The CDR (Morris, 1993) is a semi-structured interview with the patient and an informant to determine the presence and severity of dementia. It rates the subject’s cognitive performance in six domains: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. Each domain and the global CDR are rated on five levels of impairment, where zero indicates no dementia, and CDR 0.5, 1, 2, and 3 indicate questionable, mild, moderate, and severe dementia, respectively.

Memory. Rey’s Auditory Verbal Learning Test (Rey, 1964). A series of 15 one-syllable words is read five times. After each presentation participants are asked to recall which words they remember. Twenty minutes after this immediate recall (IR) condition, delayed recall (DR) is tested. Alternate forms were used at follow-up. Results were expressed as standard T-scores corrected for age, gender, and education (Schmand et al., 2012).

The Rivermead Behavioral Memory Test Logical Memory (RBMT LM) subtest (Wilson, Cockburn, & Baddely, 1985). Participants are asked to immediately reproduce (IR) two news items read to them. After 15 min DR follows. Parallel story versions were used at follow-up. Results were expressed as T-scores corrected for age, gender, and education (Schmand et al., 2012).

The Visual Association Test (VAT; Lindeboom & Schmand, 2003). Participants are shown pictures of two common objects each, representing an unusual combination, for example, an ape carrying an umbrella. Recall is tested without a delay by showing one object and asking what the other object is missing. The short version (6 items) was used for patients younger than 65 years of age and the long version (12 items) was administered to patients older than 65 years of age. The pictures are shown twice, except when recall was fully correct after the first presentation. Score is the total number of objects correctly recalled, converted to the long version with two presentations (i.e., maximum score is 24 items correct).
The enhanced cued recall test (ECR; Grober, Buschke, Crystal, Bang, & Dresner, 1988). Participants are required to recall 16 items with help of category cues. After a short distraction task, they are asked to freely recall as many of the pictures as possible; category cues are provided for the remaining items. Score is total number of objects correctly recalled.

Language. Category fluency (Luteijn & Barelds, 2004). The test consists of naming animals and occupations for 1 min each. Results were corrected for age and education (Schmand et al., 2012). Score is the mean demographically corrected T-score of the two categories.

Controlled Oral Word Association Test (Benton, Hamsher, & Sivan, 1983). During 1 min, participants are asked to say as many words as possible that begin with a given letter. Three trials with different letters were done. At follow-up, an alternate version was used. Score is the raw number correct in 3 min. Results were expressed as T-scores corrected for education (Schmand et al., 2012).

Psychomotor speed and executive functions. The Trail Making Test (TMT; Reitan, 1992). Participants are asked to connect numbers (part A) and connect numbers alternating with letters (part B). Scores are time to completion in seconds. Furthermore, a score on part B corrected for the time on part A (TMT B/A) is calculated to reflect the ability to divide attention. Results were expressed as T-scores corrected for age and education (part A) and for age, gender, and education (part B and B/A; Schmand et al., 2012).

The Stroop Color Word test (Stroop, 1935). Participants are requested to read words (card 1), name colors (card 2), and name the ink color of words (card 3) when the words are printed in a non-matching color. Scores are time to complete 100 items. Furthermore, an interference condition is calculated (card 3|2), which reflects the score on card 3 corrected for the time on card 2. Results were expressed as T-scores corrected for age and education (card 1) and for age, gender, and education (card 2, card 3, and card 3|2; Schmand et al., 2012).

The Letter Digit Substitution Test (Jolles, Houx, Van Boxtel, & Ponds, 1995). Participants must substitute over-learned signs according to a substitution key. The key shows the numbers 1–9, each paired with a different letter. The score is the number of correct substitutions made in 60 s. Results were expressed as z-scores corrected for age and education.

Visuospatial/constructive skills. The Clock Drawing Test (Royall, Cordes, & Polk, 1998). Patients are requested to draw the face of a clock on a blank piece of paper setting the hands at 1:45. The drawing is scored on a scale with a maximum of 14 points.

The subtest Block Design of the Wechsler Adult Intelligence Scale-Third edition (Wechsler, 1997). Patients must arrange colored blocks according to increasingly difficult patterns. A time bonus is given for quick solutions. Results are expressed as scaled scores corrected for age.

Assessment of symptom validity. The WMT (Green, 2003). Participants are asked to memorize a list of 20 semantically related, simple word pairs (e.g., “gold”/“silver”). After the list has been presented twice, a forced choice Immediate Recognition (IR) subtest is given, where the person should select from a new pair of words the word that was in the original list (e.g., “gold” from the word pair “gold-bronze”). Without notice, the forced choice Delayed Recognition (DR) subtest is presented after a 30-min delay in the same way as the IR, except with different distracter words. Based on the IR and DR, the consistency (CNS) score is calculated as percent agreement in responses between the two recognition trials. Any score on IR, DR, and CNS below the recommended cut-off point is indicative of non-credible performance. The effort measures are followed by three subtests of gradually increasing difficulty, which measure verbal memory ability. The first subtest is the Multiple Choice (MC) test in which the first word from each pair is shown and the corresponding word should be chosen from eight options. On the Paired Associate (PA) subtest, the investigator gives the first word and the participant is asked to name the appropriate word. In the Free Recall (FR) subtest, participants are asked to recall as many words as possible from the original list of word pairs. Auditory and visual feedback is given after each response on the IR, DR, and MC subtests. These measures, therefore, serve not only as effort subtests but also as additional learning trials, which assist motivated participants in learning. The PA subtest provides further exposure to the first words of each pair, again serving a dual purpose as a test of memory and as a learning trial, prior to FR. The effort subtests IR, DR, and CNS are relatively easy, whereas the conventional memory subtests MC, PA, and FR are relatively difficult.

Failure on one or more of the easy effort subtests was defined by the cut-off scores published in the WMT manual (Green, 2003). As explained in the “Introduction” section, the dementia profile is defined as failure on one or more of the easy effort subtests, with a difference greater than 30 percentage points between the means of the easy and difficult subtests. In this respect, it is important to note that the Advanced Interpretation program (Green, 2009) makes it clear that the dementia profiles for the WMT, MSVT, and NV-MSVT were all defined so that specificity would be very high in those with dementia. However, this means that the sensitivity to poor effort in simulators is relatively low. If a dementia profile is found in a sample where poor effort is a rare phenomenon, genuine impairment is probably the explanation of the profile. However, when the risk of poor effort is high and when actual
severe impairment is unlikely, such as in mild brain injury in compensation seekers many months post-injury, a dementia profile probably reflects poor effort. Ideally, the WMT, MSVT, and NV-MSVT will be used in conjunction with each other to optimize sensitivity to poor effort while maintaining high specificity. Furthermore, failure without a dementia profile means poor effort and unreliable results irrespective of diagnosis. To conclude that a dementia profile implies genuine severe impairment, clinical correlation is required. For example, mild head injury cannot cause failure on the easy subtests of the WMT and so even if the profile looks like a dementia profile, poor effort would be concluded.

**Group allocation: Credible versus non-credible.** Participants were divided into three groups: (1) people who passed the WMT effort subtests (WMTpass group), (2) people who failed the WMT easy subtests, but showed a dementia profile (WMTdem group), and (3) people who failed the effort subtests but did not have a dementia profile, thus showing non-credible performance (WMTnoncr group).

**Data analyses**

Statistical testing was done with SPSS for Windows (version 19.0). Group differences in age and education level were tested using a one-way analysis of variance (ANOVA) and group differences in gender were analyzed using a chi-square test. If, even after transformation, variables were non-normally distributed, non-parametric tests were performed. At baseline and follow-up, the chi-square test was used to assess the relation between the dementia profile and scores on the CDR. CDR scores were dichotomized as 0 (normal score, 0) or 1 (deviant, CDR score ≥ 0.5). Patients, who were unable to participate in follow-up due to severe dementia, were given a deviant CDR score at follow-up derived from clinical judgment (n = 16). To evaluate differences between groups on the neuropsychological tests, multivariate ANOVA (MANOVA) was used for demographically corrected standard scores. Effect sizes were calculated with partial eta squared. Post hoc analyses were performed with Bonferroni correction. A non-parametric equivalent (Kruskal–Wallis) was used to test group differences for not normally distributed variables (MMSE, ECR, VAT, Clock Drawing Test, and Block Design). Mann–Whitney tests were used for post hoc analyses. In order to obtain an estimate of the effect sizes for non-parametric analyses, univariate ANOVA was performed using Van der Waerden-normalized scores, although this transformation was not sufficiently strong for all scores to be normally distributed. Missing values at baseline (1.9% of all data points) and at follow-up (5.2% of all data points) were replaced by the group mean of each test.

**Results**

**Participants**

Three participants were excluded from further analysis due to missing WMT scores as a result of extreme slowness during the neuropsychological examination. At baseline, there were 63 patients in the WMTdem group, 92 patients in the WMTpass group and 12 participants in the WMTnoncr group. WMT subtest scores for the three patient groups are shown in Table 1. Demographic characteristics of these groups are presented in Table 2. The groups were significantly different in age, $F(2,164) = 11.50, p < .001$, where patients in the WMTdem group were older than the WMTpass and the WMTnoncr participants ($p \leq .001$ for both comparisons). The groups differed significantly in education, $F(2,164) = 3.92, p = .022$. The WMTpass group had a slightly higher educational level than the WMTdem group ($p = .038$). There were no significant group differences in gender.

At follow-up, neuropsychological test scores of 34 WMTdem patients, 67 WMTpass patients, and 8 participants with non-credible performance were available. The remaining 58 individuals did not participate in follow-up for various reasons (e.g., refused to participate, too ill, deceased). Again, there was a significant difference between groups in age, $F(2,106) = 12.60, p < .001$. Patients in the WMTdem group were significantly older than those in the WMTpass and the WMTnoncr groups ($p < .001$ and $p = .005$, respectively). There was a significant group difference in educational level, $F(2,106) = 5.99, p = .003$. Patients in the WMTpass group were better educated than WMTdem group participants ($p = .01$). No significant gender differences were found between groups.
There was a significant difference at baseline and at follow-up between the WMTdem and WMTpass groups in the CDR score (Table 3). At baseline, a much greater proportion of WMTdem patients had a deviant CDR than participants who passed the WMT ($\chi^2 = 33.02, p < .001$). At follow-up, this proportion was even larger ($\chi^2 = 35.30, p < .001$).

At baseline, the sensitivity of the dementia profile for a deviant CDR score was 60% and its specificity was 83%, not considering the non-credible group. The PPV was 81%. The dementia profile at baseline had a sensitivity of 59% and a specificity of 94% for a deviant CDR score at follow-up. The PPV was 93%.

Table 4 presents the mean scores on cognitive tests for the three patient groups at baseline and at follow-up. At baseline, significant group differences were found for all tests, except for the Stroop interference condition. Post hoc analyses showed that the WMTdem group performed significantly worse than the WMTpass group on all tests. WMTdem patients scored either worse than the WMTnoncr group or at a comparable level. For all tests with significant group differences, partial eta-squared was medium to large, with the memory tasks showing the largest effect sizes.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Follow-up</th>
<th>Post hoc analyses</th>
<th>Baseline</th>
<th>Follow-up</th>
<th>Post hoc analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WMTdem x</td>
<td>WMTpass y</td>
<td>WMTnoncr z</td>
<td>F (df)</td>
<td>WMTdem x</td>
<td>WMTpass y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MMSE raw Memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAVLT T</td>
<td>31.2 (11.7)</td>
<td>45.7 (10.4)</td>
<td>36.6 (13.1)</td>
<td>0.20***</td>
<td>25.8 (10.1)</td>
<td>43.2 (13.8)</td>
</tr>
<tr>
<td>RAVLT DR T</td>
<td>28.1 (11.9)</td>
<td>47.0 (11.3)</td>
<td>36.5 (10.7)</td>
<td>0.38***</td>
<td>22.6 (7.9)</td>
<td>43.0 (14.1)</td>
</tr>
<tr>
<td>RBMT IR T</td>
<td>35.0 (8.2)</td>
<td>46.6 (11.3)</td>
<td>44.0 (12.1)</td>
<td>0.23***</td>
<td>32.5 (8.7)</td>
<td>47.7 (13.0)</td>
</tr>
<tr>
<td>RBMT DR T</td>
<td>32.7 (9.1)</td>
<td>45.4 (11.1)</td>
<td>42.8 (11.9)</td>
<td>0.25***</td>
<td>31.1 (8.2)</td>
<td>46.0 (12.7)</td>
</tr>
<tr>
<td>ECR raw</td>
<td>11.1 (3.4)</td>
<td>14.9 (1.6)</td>
<td>14.5 (1.7)</td>
<td>0.35***</td>
<td>7.9 (4.3)</td>
<td>14.3 (2.8)</td>
</tr>
<tr>
<td>VAT raw</td>
<td>14.7 (6.8)</td>
<td>21.4 (3.8)</td>
<td>21.1 (5.2)</td>
<td>npar</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category fluency T</td>
<td>39.7 (9.2)</td>
<td>46.4 (10.4)</td>
<td>41.5 (9.6)</td>
<td>0.10***</td>
<td>33.3 (10.3)</td>
<td>45.9 (10.7)</td>
</tr>
<tr>
<td>COWAT letter fluency T</td>
<td>41.0 (8.8)</td>
<td>46.6 (10.5)</td>
<td>42.3 (10.4)</td>
<td>0.07***</td>
<td>35.6 (9.8)</td>
<td>49.0 (11.4)</td>
</tr>
<tr>
<td><strong>Psychomotor speed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop 1 T</td>
<td>39.5 (12.8)</td>
<td>45.0 (11.1)</td>
<td>29.9 (17.7)</td>
<td>0.11***</td>
<td>36.8 (12.6)</td>
<td>45.3 (11.5)</td>
</tr>
<tr>
<td>Stroop 2 T</td>
<td>37.5 (14.3)</td>
<td>46.1 (12.4)</td>
<td>29.9 (14.6)</td>
<td>0.14***</td>
<td>32.8 (13.3)</td>
<td>45.1 (13.0)</td>
</tr>
<tr>
<td>TMT A T</td>
<td>37.0 (19.5)</td>
<td>48.9 (13.7)</td>
<td>32.9 (22.8)</td>
<td>0.13***</td>
<td>25.5 (31.8)</td>
<td>50.2 (12.8)</td>
</tr>
<tr>
<td>LDST z</td>
<td>-1.1 (1.0)</td>
<td>-0.2 (1.1)</td>
<td>-1.0 (1.1)</td>
<td>0.15***</td>
<td>-1.2 (0.9)</td>
<td>-0.1 (1.2)</td>
</tr>
<tr>
<td><strong>Executive functions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop 3 T</td>
<td>37.7 (10.8)</td>
<td>44.7 (10.4)</td>
<td>33.2 (11.8)</td>
<td>0.13***</td>
<td>34.3 (7.8)</td>
<td>43.9 (12.1)</td>
</tr>
<tr>
<td>Stroop 3/2 T</td>
<td>45.1 (9.5)</td>
<td>48.0 (8.9)</td>
<td>45.6 (8.1)</td>
<td>0.02***</td>
<td>42.1 (7.3)</td>
<td>47.5 (9.7)</td>
</tr>
<tr>
<td>TMT B T</td>
<td>27.5 (19.2)</td>
<td>43.5 (16.0)</td>
<td>37.6 (18.1)</td>
<td>0.16***</td>
<td>26.6 (17.0)</td>
<td>43.9 (16.5)</td>
</tr>
<tr>
<td>TMT B/A T</td>
<td>30.5 (17.1)</td>
<td>43.4 (14.7)</td>
<td>44.7 (13.3)</td>
<td>0.14***</td>
<td>30.3 (14.2)</td>
<td>42.8 (15.3)</td>
</tr>
</tbody>
</table>

(continued on next page)
Table 4. (continued)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Follow-up</th>
<th>Post hoc analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WMTdem</td>
<td>WMTpass</td>
<td>F (df)</td>
</tr>
<tr>
<td>Visuospatial/constructive skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clock drawing test</td>
<td>9.8 (2.9)</td>
<td>11.1 (2.1)</td>
<td>11.9 (2.0)</td>
</tr>
<tr>
<td>Block Design SS</td>
<td>7.2 (2.6)</td>
<td>9.1 (3.0)</td>
<td>8.3 (3.4)</td>
</tr>
</tbody>
</table>

F = Pillai’s Trace; df = (hypothesis degrees of freedom, error degrees of freedom); ns = not significant; η² = partial eta squared. Values are expressed as mean (SD), unless otherwise indicated.

MMSE = Mini Mental State Examination; RAVLT = Rey Auditory Verbal Learning Test; IR = immediate recognition; T = normally distributed score with a mean of 50 and a standard deviation of 10, corrected for age and education; DR = delayed recognition; RBMT = Rivermead Behavioral Memory subtest story recall; ECR = Enhanced Recall; VAT = Visual Association Test; WMT = Word Memory Test; WMTdem = patients with a dementia profile on the WMT; WMTpass = people who passed the WMT effort subtests; WMTnoncr = patients with non-credible performance; npar = non-parametric.

aKruskal-Wallis.
bMANOVA.
*p < .05.
**p < .01.
***p < .001.
At follow-up, significant group differences were found for all tests. Post hoc analyses revealed a consistent pattern: patients in the WMTdem group performed significantly worse on all tests than the WMTpass group, and on the majority of tests, they also scored worse than the WMTnoncr patients. The WMTpass group outperformed the WMTnoncr participants on the ECR. For all tests, the partial eta-squared was large, with the exception of a medium size for Stroop card 3/2. In general, larger effect sizes were found at follow-up than at baseline.

**Determining Decline in Cognition for the 3 Patient Groups**

CDR decline was defined as an increase in the CDR score at follow-up compared with baseline, which implies a progression from MCI to dementia or from normal cognition to MCI (Table 5). There was a significant difference between groups at follow-up ($\chi^2 = 29.05, p < .001$). The results showed that more WMTdem patients had an increase in the CDR score than patients without the dementia profile or patients with non-credible performance.

The WMTdem group displayed a significant decline between baseline and follow-up on most tests (Table 6). The WMTpass group also performed significantly worse at follow-up compared with baseline on a number of tests, but in general they had...
declined less than the WMTdem group. For the WMTnoncr group, no significant differences were found between baseline and follow-up test scores, except for an improved performance at follow-up for Stroop interference condition and a decline in performance for Clock Drawing.

**Differences in Rate of Decline Between Groups**

Mean difference scores (the difference in test scores between baseline and follow-up) were calculated for each group (Table 6). On the tests where both the WMTpass and WMTdem groups showed cognitive decline, either a significantly greater cognitive decline was found for the WMTdem group or a trend towards such a difference in decline was visible.

**Dementia Profile and Clinical Diagnoses**

After consideration of the neuropsychological evaluation and the MRI scan made at baseline, all patients received a clinical diagnosis from the referring physician according to the criteria of the DSM-IV-TR (American Psychiatric Association, 2000). As Table 7 shows, there was a significant difference in diagnosis between groups \( \chi^2 = 57.87, p < .001 \). In the WMTdem group, a much greater proportion of patients received a diagnosis of MCI or dementia than in the WMTpass group. In the WMTnoncr group, two patients were classified as demented and one patient received at baseline a MCI diagnosis. At follow-up, the referring physician reviewed the clinical diagnosis after the follow-up neuropsychological results and comparisons with baseline test results were reported back. Using this final clinical diagnosis, results showed a significant difference between groups \( \chi^2 = 43.87, p < .001 \). Patients in the WMTdem group were more often diagnosed with MCI or dementia than patients in the WMTpass group. In the WMTnoncr group, two patients received a diagnosis of degenerative brain disease.

**Discussion**

We examined whether a dementia profile on the WMT is a significant indicator of cognitive impairment and/or cognitive decline. In addition, we evaluated the classification accuracy of the dementia profile for the clinical diagnosis. Previous reports examining WMT profiles looked the other way around (i.e., they looked at the performance of individuals with a specific neurological diagnosis).

At baseline and follow-up, significantly more patients with a dementia profile than participants who passed the WMT had a deviant CDR score. Furthermore, the WMTdem group performed significantly worse than the WMTpass group, and they declined more between baseline and follow-up on most tests. The largest effect sizes were found for memory tests. The lack of decline on the RBMT LM test is a notable exception. This is probably due to a floor effect at baseline, which left little room for deterioration.

These findings are in line with the results of a previous study that analyzed the dementia profile on the MSVT (Axelrod & Schutte, 2010). As opposed to their findings and other existing work (Howe et al., 2007; Howe & Loring, 2009), however, the percentage of patients presenting with the dementia profile in our study was not comparable with that of patients who showed non-credible performance. This inconsistency is probably due to age differences in the study samples. The age range of our study group was smaller than in the samples of previous studies. Old age is a primary risk factor for cognitive decline (Tilvis et al., 2004). Furthermore, SVT failure is higher in patients under 65 years of age (Rienstra et al., 2012). This may explain the large number of people with the dementia profile relative to the number of non-credible patients in our study. Furthermore, the Axelrod and Schutte data (2010) came from a sample of US military veterans. This population has a high rate of SVT failure, likely due in large part to the potential disability implications of evaluations done in this setting (Armistead-Jehle, 2010; Jones, Ingram, & Ben-Porath, 2012; Nelson et al., 2010).

---

**Table 7. Clinical diagnoses at baseline (provisional diagnosis) and follow-up (final diagnosis)**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Baseline</th>
<th></th>
<th></th>
<th>Follow-up</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WMTdem</td>
<td>WMTpass</td>
<td>WMTnoncr</td>
<td>WMTdem</td>
<td>WMTpass</td>
<td>WMTnoncr</td>
</tr>
<tr>
<td>Dementia</td>
<td>15 (28.8%)</td>
<td>3 (3.0%)</td>
<td>2 (1.6%)</td>
<td>19 (55.9%)</td>
<td>7 (10.5%)</td>
<td>2 (25%)</td>
</tr>
<tr>
<td>MCI</td>
<td>23 (44.2%)</td>
<td>23 (29.1%)</td>
<td>1 (8.3%)</td>
<td>8 (23.5%)</td>
<td>15 (22.4%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Worried well</td>
<td>2 (3.8%)</td>
<td>34 (43%)</td>
<td>1 (8.3%)</td>
<td>1 (2.9%)</td>
<td>35 (52.2%)</td>
<td>3 (37.5%)</td>
</tr>
<tr>
<td>Other</td>
<td>12 (23.1%)</td>
<td>19 (24.1%)</td>
<td>8 (66.7%)</td>
<td>6 (17.6%)</td>
<td>10 (14.9%)</td>
<td>3 (37.5%)</td>
</tr>
</tbody>
</table>

Baseline: 27 missing. WMT = Word Memory Test; WMTdem = patients with a dementia profile on the WMT; WMTpass = people who passed the WMT effort subtests; WMTnoncr = patients with non-credible performance; other = neurological and/or psychiatric disease other than MCI or dementia.
Regarding the ability of the WMT dementia profile to provide information about the clinical diagnosis, we investigated the classification accuracy of the profile in our sample. Sensitivity of the profile (i.e., the percentage of individuals with a deviant CDR who had a dementia profile at baseline) was a moderate 60%. However, the PPV in this memory clinic sample was high, viz. 81% at baseline and 93% at follow-up. This indicates that if a memory clinic patient has a WMT dementia profile, the clinician can be fairly confident that the patient is at least in a pre-dementia stage. These figures correspond to the results of an earlier report that evaluated the sensitivity of the MSVT dementia profile in memory clinic patients (Howe & Loring, 2009). The easy-hard criterion might also be applied to patients who do not score below the cut-offs on the symptom validity indices to provide clinically relevant information (Howe & Loring, 2009). And indeed, a vast majority of the patients in our WMTpass group, who had a deviant CDR score, showed such a large easy-hard difference (data not shown). Thus, it is likely that the sensitivity of the dementia profile would have been much higher, if we had applied the easy-hard criterion to all patients instead of to patients who failed the WMT easy subtests. Specificity of the profile (i.e., the percentage of individuals with a normal CDR who had normal effort scores on the WMT) was high at baseline (83%) and even higher at follow-up (93%).

For the non-credible performance group, we hypothesized that they would not show cognitive decline after 2 years, but might even show improved test performance. The results largely confirmed this expectation. An improved performance at follow-up was found for the Stroop interference condition, and there was only a decline in performance on the clock drawing task. Given the small sample size of the non-credible group and the large number of test variables, these differences might be due to chance fluctuations. Nevertheless, it is interesting to note that at follow-up, the WMTpass group outperformed the WMTnoncr participants on the ECR.

The majority of symptom validity techniques make use of force-choice recognition across multiple trials to detect non-credible performance (Bickart, Meyer, & Connell, 1991). However, this result suggests that paradigms requiring participants to recall relevant information (Howe & Loring, 2009). And indeed, a vast majority of the patients in our WMTpass group, who had a deviant CDR score, showed such a large easy-hard difference (data not shown). Thus, it is likely that the sensitivity of the dementia profile would have been much higher, if we had applied the easy-hard criterion to all patients instead of to patients who failed the WMT easy subtests. Specificity of the profile (i.e., the percentage of individuals with a normal CDR who had normal effort scores on the WMT) was high at baseline (83%) and even higher at follow-up (93%).

Furthermore, it is remarkable that at baseline and at follow-up, a quarter to one third of the non-credible group were given a deviant CDR by dementia specialists. However, the diagnosis of (preclinical) dementia may be invalid, since these patients were not performing to the best of their ability. These results underscore the importance of administering formal tests of symptom validity in the neuropsychological examination of patients presenting with cognitive complaints that may signify an early stage of dementia (Bush et al., 2005; Rienstra et al., 2012). Also, the CDR classification depends on subjective judgment by the clinician of the person’s everyday functioning (Lim, Chong, & Sahadevan, 2007). In the current study, however, the CDR was not the only measure used for cognitive evaluation, and accordingly, not the only measure on which we based our conclusions.

Our study has a number of strengths. First, we evaluated the dementia profile of the WMT in a large sample with a high risk of false-positive SVT results. According to Green and colleagues (2011), the main purpose of the dementia profile is “to achieve the lowest possible rate of false positives among those in whom there is a reasonable possibility of severe impairment.” For all participants in our sample, the differential diagnosis included a possible early stage of dementia at the time of referral. Second, our study is longitudinal, which is an important methodological strength when the subject is cognitive decline or dementia. Third, unlike most studies on possible pre-dementia patients, we did not highly select our patients. We did not limit our sample to unequivocal MCI patients without psychiatric co-morbidity. Instead, we only excluded clear dementia cases, so that we retained a mix of “difficult” memory clinic patients. This aspect will enhance the generalizability and relevance of our findings to clinical practice.

There are also several limitations. The first concerns the small size of the non-credible SVT group. This does not imply, however, that the influence of non-credible performance on neuropsychological test results should be underestimated. It is quite possible that the low incidence of non-credible performance in our study was caused by selection bias. Exaggeration of symptoms or impairments is probably a common cause of non-credible performance in clinical practice when patients feel the need to get recognition for their complaints (Miller, 2001). Our patients were asked to take part in a research project, which inherently implies recognition for their complaints. This may have reduced any tendency to exert non-credible performance. A second limitation is that we did not systematically explore reasons for SVT failure. This limits any attempt to explain the causes for SVT failure. Most likely, there was a mixture of reasons for failure on SVTs in the non-credible group. Yet, although the exact mechanism is important by itself, it is irrelevant to our main purpose, which is the neuropsychological characterization of patients with the WMT dementia profile. Third, because of the differences between groups in age and education, demographically corrected T-scores were used. T-scores, however, were not available for all tests. All tests for which raw scores were used were non-normally distributed, so non-parametric analysis was required. Unfortunately, with this analysis, variation in age and education was not accounted for. Nevertheless, these results were similar to the results using T-scores, so it is unlikely that age and education account for the differences between groups.

In conclusion, the results of this study demonstrate that patients with the WMT dementia profile have a high chance of showing real cognitive impairment at the moment of evaluation, and even more so 2 years later. They showed a faster cognitive decline than patients who passed the easy effort subtasks. These findings underscore the importance of administering all WMT subtasks. The first three subtests must be placed in the context of the last three to fully understand the profile. The dementia profile aids in
interpretation of the WMT when it is utilized as an SVT, and it provides clinically relevant information on (future) cognitive impairment and early dementia.

**Funding**

This study was financed by the Psychology Research Institute of the University of Amsterdam, the Departments of Neurology and Radiology of the Academic Medical Center of the University of Amsterdam, and the Graduate School of Neurosciences Amsterdam-Rotterdam.

**Conflict of Interest**

None declared.

**Acknowledgement**

The authors thank Justine Aaronson, Nina Albisser, and Hyke Tamminga for their assistance in neuropsychological testing.

**References**


