The Word Memory Test and the Validity of Neuropsychological Test Scores

Paul Green

Neurobehavioural Associates

Paul R. Lees-Haley

Independent Practice

Lyle M. Allen III

CogniSyst, Inc.

SUMMARY. The Word Memory Test (WMT; Green, Allen, & Astner, 1996) contains measures that are very sensitive to exaggeration or poor effort but insensitive to all but the most extreme forms of cognitive impairment. The WMT is unique among symptom validity tests because of its extensive validation in clinical forensic settings, rather than relying on simulation research with healthy volunteers. Effort measured by the WMT predicted 50% of the variance in a total of 30,736 neuropsychological test results from 904 consecutive patients involved in com-
pensation claims. In group data, WMT-measured effort was sufficient to eliminate or reverse major effects, such as the presence of greater impairment in people with severe versus mild head injuries. The removal of invalid data from people failing the WMT was found to make a major difference to the conclusions of several studies. Properly interpreted, the WMT will meet Daubert challenges.

KEYWORDS. Effort, malingering, brain injury, symptom validity

INTRODUCTION

When the Word Memory Test (WMT; Green, Allen, & Astner, 1996) was invented, it was already clear from a growing literature that there was a risk of concluding that a person had impairment of a certain ability, such as verbal memory, when the real reason for the low test score was deliberate exaggeration or insufficient effort to produce valid data. There was no a priori reason for assuming that misleading test results would only be found in people with a particular diagnosis, and so the WMT was given to every outpatient referred to the first author for neuropsychological assessment. Over several years, the WMT has been given to over 1,000 consecutive adults in this series, all involved in compensation or disability claims. This has become the central data set from which much of the WMT research has evolved. Each patient was given two days of assessment, including neuropsychological tests, along with various other measures of symptom validity for cross-validation. These included the Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox; 1997), the Recognition Hits and Logit Formula derived from the California Verbal Learning Test (CVLT; Delis, Kramar, Kaplan, & Ober, 1987; Millis, 1999) and Warrington’s Recognition Memory Test for Words and Faces (WRMT; Warrington, 1984). A number of patients were also given the English version of the Amsterdam Short Term Memory Test (ASTMT; Schmand, Lindeboom, Schagen, Heijt, Koene, & Hamburger, 1998) or the 21 item test (Iverson, 1998).

The main sample included 535 patients with head injuries and 89 neurological patients with strokes, aneurysms, multiple sclerosis, tumor, epilepsy or other miscellaneous conditions. Also included were patients referred with major depression (n = 85) anxiety disorders (n =
18), orthopedic injuries (n = 77), chronic fatigue syndrome (n = 34), chronic pain syndrome or fibromyalgia (n = 61) and various other conditions (n = 101). Various other groups are described below, including patients who were asked to simulate memory impairment. With data from the series of 1,000 consecutive cases, we can study not only effort itself but, more importantly, we can determine the influence of varying degrees of effort on many neuropsychological test scores and symptom rating scales. These data have revealed unexpectedly large and widespread effects of effort on test scores. It will be shown that, in the samples studied, poor effort, as measured by the WMT, had a much larger effect on neuropsychological test scores than did severe brain injury or neurological diseases.

**STRUCTURE OF THE WMT**

The WMT is a test of verbal learning and memory, designed to allow the evaluation of a person’s effort to do well in the course of taking the test, so that a determination can be made about whether or not the person’s test scores are valid estimates of ability. The WMT measures the ability to learn a list of 20 word pairs (e.g., pig-bacon, fish-fin, dog-cat) presented orally or on a computer screen. The effort components of the WMT were designed to avoid confusing actual impairment with deliberate exaggeration. They are virtually insensitive to all but the most extreme forms of impairment of learning and memory, and the range of genuine scores is very narrow, as shown below.

After being shown 20 word pairs on the computer screen, the person is required to choose the word from the original list in each of 40 new word pairs (e.g., “dog” from “dog-rabbit”). This is the immediate recognition trial (IR), the first measure of effort. Feedback regarding correctness is given to assist motivated patients in learning for later subtests. After a half-hour delay, the delayed recognition trial (DR) is presented, which is very similar to IR, but it includes different foil words (e.g., “dog-rat”). Several measures of response consistency are computed, including the consistency of responses between the 40 paired IR and DR trials, which correlates very highly with results from the IR and DR subtests.

Following the main effort measures, the patient is given four separate measures of memory ability: the Multiple Choice subtest (MC), in which the person is shown the first word from each pair and is asked to choose the matching word from eight options; Paired Associates (PA),
in which the person is given the first word from each pair by the tester and is asked to tell the tester the second word; Delayed Free Recall (DFR), in which the person is asked to recall as many words as possible from the list in any order. The Long Delayed Free Recall subtest (LDFR) may also be given after a further 20-minute delay. The computer printout shows the patient’s results in two graphs and lists the person’s raw scores on all measures of the WMT. A table shows the patient’s scores compared with the mean scores obtained from each of a number of reference groups, including patients with moderate to severe brain injury, neurological patients, patients with normal verbal memory, patients with impaired verbal memory and bright normal controls.

A cardinal feature of the WMT is that it consists of multiple subtests that vary widely in their objective difficulty level. Only one other test that measures symptom validity shares this primary characteristic, the Validity Indicator Profile (VIP; Frederick, 1997). However, unlike the VIP, the WMT was designed with the dual purpose of measuring both effort and actual ability, and the WMT measures the most common form of exaggeration, which is exaggeration of memory difficulties. Given the structure of the WMT, it is very difficult for a person who is not making a full effort to produce a plausible profile. Users of the WMT will recognize that there are many other WMT results and statistics listed in the computer printout that may be useful in the detection of exaggeration and in quantifying memory ability. The WMT can be administered orally or partially unattended on a PC-compatible computer. The two forms of the WMT appear to be equivalent, but the majority of WMT research to be discussed below has employed the computerized form.

**WMT SCORES IN PATIENTS WITH BRAIN INJURIES AND NEUROLOGICAL DISEASES**

Table 1 shows the mean scores on the WMT effort subtests from 40 healthy volunteers (Iverson, Green, & Gervais, 1999), compared with 57 patients with moderate to severe brain injuries whose mean Glasgow Coma Scale score was 9. These patients were described in detail by Allen and Green (1999) in a supplement to the WMT manual. The mean scores from the brain-injured patients on both IR and DR were above 95% correct (i.e., above 38 out of 40 correct). The mean score on the three WMT effort measures in the healthy volunteers was 97.8% correct, compared with 95.1% correct in the patients with brain injuries.
Allen and Green (1999) also showed that, within the brain-injured patients, the WMT effort measures (IR, DR and Consistency) were unrelated to major measures of head injury severity, including Glasgow Coma Scale scores, duration of post-traumatic amnesia and duration of loss of consciousness.

Further evidence that the WMT effort scores are relatively insensitive to brain disease is listed in Table 2, showing the mean scores from 40 patients with neurological diseases, including brain tumors, strokes, multiple sclerosis and ruptured cerebral aneurysms. These patients are described both individually and as a group by Green and Allen (1999). The neurological patients were divided into two groups of equal size, based on scoring above or below the groups’ median on the learning trials of the CVLT. Despite a significant difference on the CVLT in these two subgroups, there were no significant differences in their scores on the WMT effort measures nor on an independent effort measure, the CARB (Table 2). These data show that the WMT effort scores are relatively insensitive to actual cognitive impairment and, in particular, are unaffected both by neurological diseases and by memory impairment in this sample. On the other hand, the scores on the WMT subtests designed to measure memory (MC, PA, DFR and LDFR) were all significantly lower in the neurological patients with impaired CVLT scores.
than in those with normal range CVLT scores. These data show that what the latter subtests are measuring (memory) is not the same as what the effort subtests measure (effort).

There is a strong ceiling effect on the WMT effort measures. For example, on WMT IR, even neurological patients with impaired verbal memory scored a mean of almost 100% correct (96.1%, SD = 3.3), which is almost identical to the mean score from those with normal verbal memory. In marked contrast, the neurological patients with impaired memory, as defined by their CVLT results (Table 2), scored only 36.7% correct (SD = 10.9) on the WMT free recall subtest (WMT-DFR). WMT effort scores are not normally distributed, but the WMT memory scores are normal.

<table>
<thead>
<tr>
<th>TABLE 2. No Difference in Mean Scores on CARB and the WMT Effort Measures Between Neurological Patients with “Normal Range” versus “Impaired” Memory Scores on the CVLT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effort Tests</strong></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>WMT-IR</td>
</tr>
<tr>
<td>WMT-DR</td>
</tr>
<tr>
<td>WMT-CONS</td>
</tr>
<tr>
<td>CARB total</td>
</tr>
<tr>
<td><strong>Memory Tests</strong></td>
</tr>
<tr>
<td>CVLT 1 to 5</td>
</tr>
<tr>
<td>CVLT-SD</td>
</tr>
<tr>
<td>CVLT-LD</td>
</tr>
<tr>
<td>Recog.hits</td>
</tr>
<tr>
<td>WMT-MC</td>
</tr>
<tr>
<td>WMT-PA</td>
</tr>
<tr>
<td>WMT-DFR</td>
</tr>
<tr>
<td>WMT-LDFR</td>
</tr>
<tr>
<td>VIQ</td>
</tr>
<tr>
<td>Educ (yrs)</td>
</tr>
</tbody>
</table>
SEVERITY OF HEAD INJURY

We might expect WMT effort scores to decline as the severity of brain injury increases. Paradoxically, the opposite pattern was found in a large sample of head injury patients of mixed severity (Green, Iverson, & Allen, 1999). On average, those with the most mild head injuries obtained significantly lower WMT effort scores than the patients with the most severe brain injuries. Such a result is the exact opposite of what would be expected if we were using a test sensitive to brain injury and if all test results were valid. It was argued that, as a group, the patients with mild head injuries were more inclined to exaggerate their impairment. This was supported by significantly lower scores in the mild head injury group on an independent test of effort, the CARB. When cases failing CARB were excluded, there were no longer any significant differences between the mild and the severe brain injury patients on any of the WMT effort scores. This provided more evidence that the WMT effort scores are sensitive to effort but insensitive to brain injury in the vast majority of outpatients studied.

ABNORMAL BRAIN SCAN RESULTS
AND EFFORT SCORES

In the current series of 1,000 patients, there were 187 patients referred for assessment after a head injury who had abnormal results on CT or MRI scans of the brain and 149 who had normal brain scan results. The mean Glasgow Coma Scale scores in these two groups were, respectively, 10.3 (SD = 4.2) and 14.7 (SD = 1.2). Also, the median duration of post traumatic amnesia was 48 hours in those with abnormal scans, but it was zero in the group with normal brain scans. The mean WMT-DR score in those with abnormal brain scans (90.7% correct, SD = 13) was found to be paradoxically significantly higher than the mean WMT-DR score in those with normal brain scans (82.5% correct, SD = 18.4; Mann-Whitney U test, z = 4.3, p < .0005). This difference cannot be explained as a result of more severe cognitive impairment in one group than the other, because the results are directly contrary to this hypothesis. Instead, it is more likely that, on average, symptom exaggeration was greater in those with normal brain scans than those with abnormal brain scans in this sample of patients assessed in the course of compensation claims. Indeed, on the Memory Complaints Inventory (MCI; Green & Allen, 1997), it was found that those with normal brain...
scans complained of significantly more memory problems than those with abnormal brain scans (mean MCI scores of 36.3%, SD = 17 versus 25.0%, SD = 17, F [2,300] = 9.9, p < .0005). In further support of the effort hypothesis, the mean score on an independent effort test, the CARB, was significantly higher in those with abnormal brain scans (mean 96.1% correct, SD = 9.4) than in those with normal scans (mean 91.1%, SD = 15.3; z = 3.3, p = .001). Higher WMT and CARB scores in those with the most severe brain injuries are most reasonably explained by the more severely injured patients making a greater effort to do well on these tests. The difference in scores obviously does not reflect differences in brain functioning between these two groups.

**SCORES OF VOLUNTEER SIMULATORS, PATIENTS AND PATIENT SIMULATORS ON WMT**

Table 3 shows the scores from 25 highly educated volunteers, mainly psychologists and physicians, who were asked to take the WMT and to act as if they had impaired memory. These simulators were told that the WMT was designed to detect exaggeration of impairment, and they were asked to try to circumvent the WMT effort measures and respond as if they had genuine memory impairment. Whereas 40 healthy volunteers obtained a mean score of 97.8% correct on the three WMT effort measures, the sophisticated simulators obtained a mean score of only 67.2% correct (Iverson, Green, & Gervais, 1999). On the delayed recognition subtest (DR), the simulators obtained a mean score which was 13 standard deviations lower than the mean from the healthy volunteers who made an effort to score as well as they could (Table 3). Even though these data are not normally distributed, such figures give a clear indication of the magnitude of the difference between full effort versus incomplete effort in group data. The mean WMT-DR score of the sophisticated simulators was more than five standard deviations lower than the mean score from neurological patients known to have impaired verbal memory (Table 2) and more than seven standard deviations lower than the mean score from 57 patients with a mean Glasgow Coma Score of 9 (Table 1). Thus, volunteers who were trying to act as if they had impairment and who said that they tried hard to be subtle to avoid being detected actually scored many standard deviations below people with brain damage or brain disease. Clearly, they were not very successful in their goal.

Similar results were obtained from 20 consecutive patients tested
clinically who passed the WMT effort measures during their regular disability examination. Their mean WMT-DR score was 98.2% correct ($SD = 3$), and so it was obviously very easy for them. They were asked to take the WMT again, but this time to fake memory impairment and make every effort to avoid being detected as exaggerating (Table 3). On retesting, their mean WMT effort score dropped from 98.2% to 62.6% correct. Their mean WMT-DR score of 62% was (a) more than 12 standard deviations lower than they were actually capable of achieving, as shown by their scores when they took the test the first time, and (b) 8.7 standard deviations lower than the mean score from patients with moderate to severe brain injuries. In using standard deviations with such data, there is no assumption of normality, and WMT effort test data are definitely not normally distributed. Whether data are normal or not, Tchebycheff’s theorem (Hays, 1963) states that the maximum probability of a score in a given distribution that is $k$ standard deviations from the mean is $1/k^2$ squared. Hence, in the patients with moderate to severe brain injuries in Table 1, a WMT-DR score of 62% is highly improbable (one in 76 or $p = 0.013$), and yet this was the mean score from patients trying to simulate impaired memory. Their WMT-DR scores were so low that it was highly unlikely that they were drawn from the same group as patients with brain injury.

TABLE 3. Mean Scores (Percent Correct) and Standard Deviations on Each of the WMT Effort and Memory Scores from All Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>WMT-IR</th>
<th>WMT-DR</th>
<th>Consistency</th>
<th>Multiple Choice</th>
<th>Paired Associates</th>
<th>Delayed Free Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Community volunteers</td>
<td>40</td>
<td>98.0 (2.8)</td>
<td>98.6 (2.4)</td>
<td>96.8 (3.8)</td>
<td>95.4 (6.7)</td>
<td>92.6 (9.6)</td>
<td>63.7 (12.4)</td>
</tr>
<tr>
<td>2) Sophisticated simulators</td>
<td>25</td>
<td>70.9 (12.6)</td>
<td>67.2 (16.5)</td>
<td>63.5 (12.4)</td>
<td>46.8 (17.9)</td>
<td>48.2 (19.9)</td>
<td>35.0 (19.2)</td>
</tr>
<tr>
<td>3A) *Clinical cases making genuine effort</td>
<td>20</td>
<td>98.1 (3.1)</td>
<td>98.2 (3.0)</td>
<td>96.6 (3.8)</td>
<td>94.0 (8.0)</td>
<td>92.7 (10.1)</td>
<td>54.1 (13.7)</td>
</tr>
<tr>
<td>3B) Clinical cases asked to simulate</td>
<td>20</td>
<td>63.2 (12.6)</td>
<td>62.0 (12.0)</td>
<td>62.7 (8.9)</td>
<td>39.5 (14.3)</td>
<td>40.7 (17.1)</td>
<td>21.6 (9.6)</td>
</tr>
<tr>
<td>4)**Clinical cases exaggerating</td>
<td>20</td>
<td>63.5 (18.0)</td>
<td>62.8 (22.1)</td>
<td>62.4 (12.4)</td>
<td>46.3 (20.6)</td>
<td>38.0 (15.3)</td>
<td>26.2 (10.0)</td>
</tr>
</tbody>
</table>

* Group 3A scored significantly higher than groups 3B and 4 on all WMT measures. ** Group 4 was not significantly different from group 3B on any WMT measure.
It is very likely that the exaggeration shown by the patients who volunteered to simulate impairment on the WMT is similar to what would be exhibited by other patients who spontaneously exaggerate impairment. To examine this suggestion, a group of 20 patients who spontaneously failed the WMT effort measures on clinical testing was selected from a database with results from 1160 patients tested independently by Dr. Roger Gervais for disability. These 20 patients were selected to match the patients from group 3A in Table 3 in terms of diagnosis, age, gender and years of education. Their mean score on the WMT effort measures was 62.9% correct (Table 3), which is almost identical to the mean score from the patients who were asked to fake impairment without being detected (62.6% correct). This experiment demonstrates that, on average, patients who try to present themselves as having impaired memory on the WMT, when they are not really impaired, score in the region of 5 to 10 standard deviations below people known to have actual brain injury or disease. The experimental simulating patients, the matched patients who failed the WMT spontaneously, and the sophisticated simulators all produced mean WMT effort scores more than ten standard deviations below what they were actually capable of achieving.

In 949 consecutive patients of mixed diagnoses who took the computerized WMT, the 660 cases who passed the effort measures scored a mean of 96.7% correct ($SD = 3.7$) on IR, 96.8% correct ($SD = 3.6$) on DR and 94.7% ($SD = 4.8$) on Consistency. In contrast, for the 289 patients who failed any one of the computerized WMT effort measures, the mean scores on the same measures were 73.3% correct ($SD = 15.8$) on IR, 70.4% correct ($SD = 16.5$) on DR and 68.4% ($SD = 11.5$) on Consistency. On DR, for example, those who failed the WMT scored more than seven standard deviations lower than those who passed. Such enormous differences between the mean test scores of large patient groups are very unusual indeed in psychology. It may also be noted from the standard deviations that the variation in scores in those making inadequate effort greatly exceeds the variation in those making a good effort. Almost anyone making an effort to do so can score close to 100% correct on WMT effort measures, but the range of scores in exaggerators is wide.

**EFFORT HAS A LARGER EFFECT ON NEUROPSYCHOLOGICAL TEST SCORES THAN SEVERE BRAIN INJURY**

It is essential in validating effort tests to prove that, when the test is failed, the person’s other test results are probably invalid. This step has
not been taken in the validation of most effort tests, but much of the re-
search with the WMT has been directed towards this fundamental goal.
The results provide some of the strongest evidence that failure on the
WMT implies invalid results on other tests. A hint of the likely effect of
exaggeration of cognitive deficits on neuropsychological test results in
general may be found in the mean memory test scores shown in Tables
1, 2 and 3. We know that the clinical cases discussed earlier, who were
asked to simulate impairment, did not have impaired memory because
they took the WMT and scored a mean of 54.1% correct ($SD = 13.7$) on
delayed free recall of the word list. However, when they were later
asked to exaggerate and simulate memory difficulties under experimen-
tal instructions, their mean score was 21.6% ($SD = 9.6$), which was
more than thirty points (or 2.4 standard deviations) lower than their ac-
tual ability, as previously measured. This score was lower than the
mean score from neurological patients with impaired memory (36.7%,
Table 2), and it was less than half of the mean score of 45.4% correct,
obtained from patients with head injuries sufficient to produce a mean
Glasgow Coma Scale score of 9 (Table 1). In the 1,000 case series, the
289 patients who failed the computerized WMT effort measures obtained
a mean free recall score (DFR) of 28.9% ($SD = 12.6$), which is also lower
than the mean scores from patients with brain injuries or neurological dis-
eases shown in Tables 1 and 2. These data provide one illustration of an
important general finding in these groups of patients, which is that poor
effort or exaggeration of cognitive difficulties has a larger effect on mem-
ory test scores than brain injury or neurological disease.

It is necessary to ask whether effort, as measured by the WMT, also
has a pronounced effect on scores from other tests, which are in wide-
spread use, such as the CVLT. One way of answering such a question is
to examine the mean scores on an isolated CVLT measure, such as long
delayed free recall, in people who score in all ranges on the WMT effort
measures. In the 1,000 patients tested consecutively, 949 cases took
both the CVLT and the computerized or oral WMT, and the correlation
between WMT-DR and CVLT long delayed free recall was 0.56. It is
very clear in Table 4 that effort, as measured by the WMT and as re-
lected in the mean WMT-DR score, is a continuum. The lower their
WMT effort score, the lower the patients scored on CVLT long delayed
free recall. In the group scoring 67% on WMT-DR (which is fairly typi-
cal of those making incomplete effort, irrespective of diagnosis) the
long delayed free recall score was an average of 1.2 standard deviations
lower than in those with a mean WMT-DR score of 97%, which is typi-
cal of good effort. Such large effects are not only found with the CVLT
but with almost any test so far examined.
To examine the effects of effort on neuropsychological test scores in general, the scores from the widely used tests listed in Table 5 were studied, using data from 904 clinical patients from the present series (Green, Rohling, Lees-Haley, & Allen, 2001; Rohling, Green, Allen, & Lees-Haley, 2000). All of the measures were converted to z-scores relative to external normative groups and then averaged for each case. Using this technique, the mean score from normal controls would be a z-score of zero, and non-zero scores would represent standard deviations above and below the normal mean. A composite mean score, the overall test battery mean (OTBM), was calculated from the scores from up to 43 separate measures of ability, with a mean of 34 measures for each of 904 cases, yielding 30,736 ability test scores (Table 5). In these data, the mean score from the WMT effort measures was found to explain 49% of the variance in the OTBM ($r = 0.7$). In contrast, age and education, which are widely regarded as significant determinants of test scores, explained only 4% and 11% of the variance, respectively, indicating relatively small effects of age and education on the same OTBM data, when compared with the effects of effort.

### EFFORT AS A CONFOUNDING VARIABLE IN GROUP STUDIES

Patients with mild head injuries who passed the WMT-DR effort measure produced a slightly below average OTBM mean z-score of

<table>
<thead>
<tr>
<th>Number of cases</th>
<th>Mean WMT-DR % correct (SD)</th>
<th>Mean CVLT LD FREE score out of 16 (SD)</th>
<th>LD FREE SD units below top scoring group</th>
</tr>
</thead>
<tbody>
<tr>
<td>624</td>
<td>97.6% (2.5)</td>
<td>11.0 (3.1)</td>
<td>0.0</td>
</tr>
<tr>
<td>136</td>
<td>87.5% (2.7)</td>
<td>8.9 (3.2)</td>
<td>–0.68 SD</td>
</tr>
<tr>
<td>76</td>
<td>77.1% (2.8)</td>
<td>8.0 (3.5)</td>
<td>–0.97 SD</td>
</tr>
<tr>
<td>49</td>
<td>67.6% (2.7)</td>
<td>7.2 (3.3)</td>
<td>–1.2 SD</td>
</tr>
<tr>
<td>38</td>
<td>56.7% (2.6)</td>
<td>5.1 (2.7)</td>
<td>–1.9 SD</td>
</tr>
<tr>
<td>46</td>
<td>42.0% (8.4)</td>
<td>3.9 (2.9)</td>
<td>–2.3 SD</td>
</tr>
</tbody>
</table>

Note: WMT-DR = Word Memory Test Delayed Recognition subtest; LDFREE = Long Delayed Free Recall score on the California Verbal Learning Test.
That is, on average, they scored only 0.12 standard deviations below the normal mean on a wide range of neuropsychological tests, which was not significantly different from the mean score from people with orthopedic injuries and no head injury. In contrast, those with mild head injuries who failed WMT-DR produced a mean score of $-1.36$ on the OTBM. Thus, exaggeration, as shown by low WMT effort scores, caused the mean OTBM from the exaggerators to be more than 10 times further below the normal mean than scores from patients with similar mild head injuries but who were making a genuine effort. Also, patients with severe brain injuries who passed the WMT effort tests scored only 0.39 standard deviations below the normal mean on the OTBM. Hence, the effect of exaggeration in people with mild head injuries was to suppress their overall mean performance on an average of 34 tests to a degree which was more than three times greater than the effects of a severe brain injury ($-1.36$ versus $-0.39$).

If effort has such large effects, it could easily obscure real differences between groups. In fact, there was no significant difference between the mean OTBM score in patients with mild head injuries versus those with severe brain injuries or neurological diseases (Green, Rohling, Lees-Haley, & Allen, 2001; Rohling et al., 2000). If taken out of context, such results would suggest that many of the neuropsychological tests in com-

**TABLE 5. Forty-Three Ability Measures Contributing to the Overall Test Battery Mean (OTBM-43), Grouped by Domain**

| Executive Functioning, EF (n = 6) | Wisconsin Card Sorting Test–Categories achieved & Perseverative errors; Category Test–Errors; Thurstone Word Fluency; Ruff Figural Fluency–Total score and perseverations; Gorham’s Proverbs |
| Memory and Learning, ML (n = 15) | CVLT–Total, Trial 5, SDFR, LDFR, Recognition hits; Cognisyst Story Recall Test–Immediate & Delayed Recall; Word Memory Test–Paired Associates, Multiple Choice, Delayed recall, Long delayed recall; Rey CFT–Delayed Recall & Recognition; Warrington Recognition Memory Test–Words & Faces |
| Verbal Comprehension, VC (n = 4) | WAIS-R VIQ or MAB VIQ; WRAT-3–Reading, Spelling & Arithmetic |
| Attention and Working Memory, AW (n = 8) | Trail-Making Test–Parts A & B; Digit Span–Forward & Backward; Visual Memory Span–Forward & Backward; CVLT–Trial 1 & List B |
| Perceptual Organization, PO (n = 4) | Rey Complex Figure Test–Copy & Recall; Benton’s Judgment of Line Orientation; WAIS-R PIQ |
mon use are insensitive to the effects of brain injury or disease. However, when the data from patients failing the WMT were excluded because they were assumed to be invalid, it was found that the patients with severe brain injuries and neurological diseases scored significantly lower on the OTBM than patients with mild head injuries, as we would expect. In this study, we could only see that severe brain injury had more of an effect on test scores than mild head injury once we removed the invalid data created by people failing the WMT effort tests.

**RANGES OF EFFORT AND OTBM SCORES**

Failure on WMT-IR and DR means that a person has scored about three standard deviations or more below the mean score obtained by patients with moderate to severe brain injuries shown in Table 1. We can see in Table 6 that 374 people who scored at or above the mean WMT score from these patients with brain injuries obtained a mean OTBM score of exactly zero, meaning that they performed at the same level as normal controls on the mean of 34 test scores. However, the mean OTBM was $-0.44$ in the 271 cases who scored between 0 and 2 standard deviations below the severe brain injury mean on WMT effort measures. Scores in this range would not fall below the usual (conservative) cut-offs for incomplete effort. Those who scored more than six standard deviations below the mean from people with brain injuries on the WMT effort measures ($n = 104$) obtained a mean OTBM of $-1.8$ (Table 6). Thus, their mean score across an average of 34 measures was 1.8 standard deviations below the normal mean, which represents an extremely low level of performance, when we consider that the effect of a severe brain injury was only to produce a mean OTBM of $-0.39$. In this instance, poor effort led to a mean OTBM that was 4.6 times further below the normal mean on the OTBM than the mean score of patients with severe brain injuries. Again, incomplete effort had a far greater effect on the overall test battery mean than severe brain injury. This is all the more significant because more than half of all patients in the group with a mean OTBM of $-1.8$ were cases of mild head injury, and only 14 were cases of severe brain injury or neurological disease. Therefore, it is not plausible that the mean OTBM score of $-1.8$ is a valid reflection of the patients’ abilities. These data show in another way how powerfully effort, as measured by the WMT, influences neuropsychological test scores.


In principle, effort might affect some ability tests more than others. In the study discussed above, each of the 43 test measures was assigned a priori to one of six domains of cognitive ability (see Table 5). For example, the Category Test was placed into the “executive function” domain, and others, such as CVLT scores, went into the “memory and learning” domain. Other tests were classified as measures of psychomotor skills, verbal comprehension, attention and working memory, or perceptual-organizational abilities. Within each domain, all available z-scores were averaged, yielding six domain scores per patient.

In Figure 1, the mean OTBM and domain scores are shown for five groups, broken down into the ranges within which people scored on the mean of the three main WMT effort measures. In the group on the far left, the patients’ mean WMT effort z-scores were at or above the mean for brain-injured patients. The next group consists of people whose WMT z-scores were up to two standard deviations lower than the mean from brain injured patients. Similarly, scores which were three standard deviations below the brain-injured mean would be in the third range from the left in Figure 1, and those scoring more than six standard deviations lower than people with brain injuries would fall into the last range (far right), representing very poor effort indeed. The numbers of patients scoring in each range on the WMT measures shown in the figure are as follows:

**TABLE 6. Means and Standard Deviations for up to 43 Neuropsychological Test Scores Expressed as z-Scores Relative to the Normal Mean (OTBM*)**

<table>
<thead>
<tr>
<th>WMT range relative to TBI group</th>
<th>Number of cases</th>
<th>Mean OTBM score</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above TBI mean</td>
<td>374</td>
<td>0.0</td>
<td>0.6</td>
</tr>
<tr>
<td>Mean to –2 SD</td>
<td>271</td>
<td>-0.44</td>
<td>0.6</td>
</tr>
<tr>
<td>-2 SD to –4 SD</td>
<td>87</td>
<td>-0.97</td>
<td>0.6</td>
</tr>
<tr>
<td>-4 SD to –6 SD</td>
<td>60</td>
<td>-1.34</td>
<td>1.1</td>
</tr>
<tr>
<td>Less than –6 SD</td>
<td>104</td>
<td>-1.8</td>
<td>0.9</td>
</tr>
</tbody>
</table>

*OTBM = overall test battery mean derived from the mean of 43 test z-scores.
Note: Patients are grouped according to how they scored on the WMT compared with patients with moderate to severe traumatic brain injury (TBI). Ranges of WMT scores are in standard deviations from the mean from the latter patients.

**EFFECTS OF EFFORT ON SPECIFIC COGNITIVE DOMAINS**

In principle, effort might affect some ability tests more than others. In the study discussed above, each of the 43 test measures was assigned a priori to one of six domains of cognitive ability (see Table 5). For example, the Category Test was placed into the “executive function” domain, and others, such as CVLT scores, went into the “memory and learning” domain. Other tests were classified as measures of psychomotor skills, verbal comprehension, attention and working memory, or perceptual-organizational abilities. Within each domain, all available z-scores were averaged, yielding six domain scores per patient.

In Figure 1, the mean OTBM and domain scores are shown for five groups, broken down into the ranges within which people scored on the mean of the three main WMT effort measures. In the group on the far left, the patients’ mean WMT effort z-scores were at or above the mean for brain-injured patients. The next group consists of people whose WMT z-scores were up to two standard deviations lower than the mean from brain injured patients. Similarly, scores which were three standard deviations below the brain-injured mean would be in the third range from the left in Figure 1, and those scoring more than six standard deviations lower than people with brain injuries would fall into the last range (far right), representing very poor effort indeed. The numbers of patients scoring in each range on the WMT measures shown in the figure are as follows:

**TABLE 6. Means and Standard Deviations for up to 43 Neuropsychological Test Scores Expressed as z-Scores Relative to the Normal Mean (OTBM*)**

<table>
<thead>
<tr>
<th>WMT range relative to TBI group</th>
<th>Number of cases</th>
<th>Mean OTBM score</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above TBI mean</td>
<td>374</td>
<td>0.0</td>
<td>0.6</td>
</tr>
<tr>
<td>Mean to –2 SD</td>
<td>271</td>
<td>-0.44</td>
<td>0.6</td>
</tr>
<tr>
<td>-2 SD to –4 SD</td>
<td>87</td>
<td>-0.97</td>
<td>0.6</td>
</tr>
<tr>
<td>-4 SD to –6 SD</td>
<td>60</td>
<td>-1.34</td>
<td>1.1</td>
</tr>
<tr>
<td>Less than –6 SD</td>
<td>104</td>
<td>-1.8</td>
<td>0.9</td>
</tr>
</tbody>
</table>

*OTBM = overall test battery mean derived from the mean of 43 test z-scores.
Note: Patients are grouped according to how they scored on the WMT compared with patients with moderate to severe traumatic brain injury (TBI). Ranges of WMT scores are in standard deviations from the mean from the latter patients.
ure, from the highest to the lowest ranges of WMT scores, were 374, 271, 87, 60 and 104.

Figure 1 shows that effort, as measured by the WMT, has a powerful effect on test scores in all domains, but especially on the mean from 15 test scores in the domain of memory and learning (ML-15). Although the memory and learning results included the WMT ability measures, separate analyses excluding these subtests produced very similar results to those shown in Figure 1. In the current study, when people chose to exaggerate cognitive difficulties, they were most likely to exaggerate on memory tests. In 694 patients who passed WMT-DR, the mean z-score on the memory and learning domain was $0.23$, compared with $-1.88$ in 188 cases who failed the DR effort measure. Figure 1 also shows that effort, as measured by the WMT, is a continuum, and that the lower that people score on the WMT effort measures, the lower they score across all domains. Hence, effort cannot be fully represented as two categories, which are “good effort” or “bad effort,” because it is a continuum. It is a matter of degree.

A question might be raised about the representativeness of the main 1,000 patient sample used in the WMT research. The rates of exaggeration in forensic cases estimated by Larrabee (2000) suggest that the
28.9% failure rate on the WMT in the current series of 1,000 cases is no higher than and probably lower than average for neuropsychologists in private practice. Secondly, in a very large independent series of patients with brain injury, Rohling (personal communication) has found that, using effort measures other than WMT, it was possible to explain 50% of the variance in a battery of neuropsychological test results, strongly supporting the findings with the WMT.

**OLFACTORY TEST SCORES, DEPRESSION, FIBROMYALGIA, CHRONIC PAIN AND EFFORT**

In the above examples, the effects on neuropsychological test scores of exaggeration, measured by the WMT, were so large that they literally inverted or obscured known effects. A similar phenomenon was found in studies of olfactory discrimination. Green and Iverson (1997, 2001) and Green, Rohling, and Iverson (in press) showed that impaired smell test scores were far more common in patients with the most severe brain injuries than in those with the least severe head injuries. However, exaggeration of cognitive difficulties, as shown by failing the WMT, was strongly associated with exaggeration of impairment on the smell test. The excess of impaired olfaction in those with more severe brain injuries was found only after removing patients who failed the WMT effort measures and whose data were, therefore, of questionable reliability. Hence, in studies of olfaction in groups of patients, it is important to measure effort and to isolate and remove test scores whose validity is questionable because of invalid effort. Otherwise, false conclusions may be drawn from group data.

In a meta-analysis of many studies of depressed patients, Veiel (1997) showed that test scores, such as those on Trail Making, were significantly lower than normal in all depressed patient groups. In contrast, Rohling, Green, Allen, and Iverson (in press) showed that, after removing cases failing effort tests (either the WMT or CARB), leaving 420 patients in the sample, there was no association between self-reported symptoms of depression and any of numerous neuropsychological test scores, including finger tapping speed, grip strength, or Grooved Pegboard scores. While those with very high and very low Beck Depression Inventory scores differed on other self rating scales measuring depression, none of the 43 separate t-tests showed significant differences on the neuropsychological tests employed (p > 0.05 for all tests, uncorrected). If depression did not affect scores on relatively difficult neuro-
psychological tests, it was reasoned that depression could not explain failure on effort tests. This argument was recently given added weight by a study of depression and the Test of Memory Malingering ([TOMM; Tombaugh, 1996]; Rees, Tombaugh, & Boulay, 2001).

Other studies have shown that incomplete effort is a factor in many different diagnostic conditions, including whiplash (Schmand et al., 1998), chronic fatigue syndrome (Van der Werf, Prins, Jongen, van der Meer, & Bleijenbet, 2000) and fibromyalgia (Gervais, Russell, Green, Allen, Ferrari, & Pieschl, 2001). The latter authors found only a 4% rate of failure on the WMT or CARB in 50 fibromyalgia patients who were not making a disability claim. However, there was a 35% failure rate in fibromyalgia patients claiming disability or already receiving compensation. The vast majority (82%) of the fibromyalgia patients made a valid effort on both CARB and the WMT, and so the results do not suggest that symptom exaggeration on cognitive tests is strongly associated with fibromyalgia, in particular. Instead, the study suggests that symptom exaggeration is associated with the availability of financial incentives for disability. However, even in children with no financial incentive, it has been found that effort is often insufficient to produce valid test results (Flaro, Green, & Allen, 2000). It would not come as a surprise to most teachers that, when taking tests, some children “could do better.”

In another study, Gervais, Green, Allen, and Iverson (2001) showed that there was approximately a 42% failure rate on CARB and the WMT effort tests in a subgroup of patients with pain disorder. However, when the next series of patients was warned that CARB failure is not caused by pain or emotional distress, the failure rate on CARB dropped to 6%, whereas on the WMT, about which they were not warned, the failure rate remained unchanged. It was concluded that neither pain nor emotional distress could explain the patients’ failure on the effort tests.

**SENSITIVITY OF WMT**

For clinical purposes, it is critical to assess both the sensitivity and specificity of the WMT in the identification of suboptimal performance. In the clinical patients asked to simulate impairment (Table 3), the sensitivity of the WMT to exaggeration was 100%, because all patients failed one or more of the effort measures. In the highly educated simulators, the sensitivity of the WMT was 96%, because all cases failed one or more of the effort measures but one person passed all the WMT effort measures. This subject also scored within the normal range on all mem-
ory subtests of the WMT, and so she did not present an impaired memory profile. However, her scores were probably lower than she could have obtained if trying her best. Combining these two groups of simulators, the sensitivity of the WMT to exaggeration would be 97.7%. None of the healthy volunteer controls failed any of the WMT effort measures, giving a specificity of 100% in this series. These are probably overestimates of the sensitivity of the test under realistic clinical conditions. Studies of healthy subjects making an effort or people simulating impairment experimentally are limited in terms of their generalizability to clinical cases, which is one reason that most of the WMT research has been carried out in a clinical setting with actual patients evaluated for compensation-related disability claims. In order to fail the main WMT effort measures (IR and DR), a person has to score about three standard deviations or more below people with severe brain injuries or neurological diseases. For this reason, we would expect the false positive rate to be very low.

One feature of people who exaggerate their cognitive difficulties is that they tend to make a poor effort on some tests but not on others. They will often pass one effort test but fail another one of objectively equal difficulty. They perform inconsistently, which is one reason why the WMT contains measures of consistency of performance. For example, out of 889 patients in the current series who were given both the WMT and CARB, 140 cases failed CARB, from which we may conclude that they were not producing reliable results. Of these patients, 110 cases failed the WMT, such that, if CARB were used to define exaggeration, the sensitivity of the WMT would be 78%. The specificity of the WMT as a predictor of CARB failure would be 83%, because 130 patients passed CARB but failed the WMT. Similarly, 44 patients from the 1,000 case series were given both the WMT and the ASTMT. There was agreement between these two tests about whether effort was adequate or inadequate in 86.4% of cases, but this means that more than 15% of the sample passed one of these effort tests but failed the other. Of those 10 cases who failed the ASTMT, 8 also failed the WMT. Of those 12 cases who failed the WMT, 8 also failed the ASTMT. It may be argued that, because tests like the WMT, TOMM, ASTMT and CARB are equally easy, inconsistent scores across these tests provide additional evidence of suboptimal effort and unreliable data.

The third author compiled a total of 694 patients from three different practices who were given the TOMM as well as the WMT. Seventy patients (10.1%) failed the TOMM and, of these, 67 cases failed the WMT, such that the sensitivity of the WMT to poor effort as defined by
TOMM failure was 96%. However, there were 111 cases who passed the TOMM but failed the WMT, or an additional 15.5% of the sample. The total WMT failure rate was over 2.5 times the failure rate for the TOMM. Hence, if the WMT were used to define incomplete effort, the TOMM would have a sensitivity of only 38.7% relative to the WMT. Green, Berendt, Mandell, and Allen (2000) studied 144 patients given the TOMM and the WMT. Even though none of them carried diagnoses associated with cerebral dysfunction, the patients who failed the WMT but who passed the TOMM scored significantly lower on the verbal memory subtests of the WMT than patients with severe brain injury. It was argued that these test results were implausible and a result of suboptimal performance. Why else should they score significantly worse than people known to have severe brain injuries, when they had no cerebral dysfunction? It is often stated that there is no gold standard for exaggeration. The nearest we have to a gold standard is that people with mild head injuries or diseases implying no cerebral dysfunction are not expected to score worse than people with severe brain injuries, and, if they do, their results are of doubtful validity. Such is the case with people who pass the TOMM and fail the WMT. Gervais et al. (1999) argued that more people in their study failed the WMT than the TOMM because the WMT is a verbal memory task, whereas the TOMM is based on recognition memory for pictures. Consistent with this reasoning, patients were far more likely to complain of verbal than visual-spatial memory difficulties on a structured self-report measure of memory complaints (Green & Allen, 1996, 2000). These findings were consistent across all seven diagnostic groups in the sample, and were true of all groups whether they passed or failed any of the symptom validity measures used.

Green (2001) discusses many reasons why clinicians are prone to disagreement about the validity of test results. One reason is that effort tests using different types of stimuli (e.g., verbal versus visual-spatial) are unlikely to be equivalent to each other. The WMT produced very different results than the “Tenhula-Sweet” malingering formula based on how a person performs on the Category Test (Williamson, Green, Allen, & Rohling, 2000). More studies are needed to compare different methods for measuring effort with each other and to study their relationship with symptom reporting in different clinical populations. In comparing different effort measures, it will be helpful to ask how much variance in test scores, especially memory test scores, is explained by each effort measure. It may be hypothesized that the more sensitive a
test is to variations in effort, the greater will be the variance in neuro-
psychological test scores explained by scores on the effort test.

FALSE POSITIVES IN CLINICAL INTERPRETATION

It is important to point out that failure on the WMT effort measures
does not mean that a person will automatically be labeled as a malin-
gerer. This is such a pejorative and emotionally laden term that most
medical practitioners are extremely reluctant to use it. It may be most
prudent to report WMT findings operationally and to keep any discus-
sion of motivation separate from WMT results and the validity of other
results. If a person fails the WMT effort measures, they fall into one of
two categories. In most cases, the other ability test results obtained from
that person at about the same time as the WMT testing will be invalid,
because the person probably did not make sufficient effort to produce
valid test results. It is likely that the person’s test results will overesti-
mate any impairment actually present. Alternatively, as with other ef-
fort measures, in relatively few cases, it may be concluded that the
person was actually unable to pass the WMT effort measures because of
very severe and widespread cognitive impairment.

The latter cases are quite rare and should be obvious to the clinician.
In the 1,000 case series described above, it is thought that there were
only about five cases of this type, and they can be used as a benchmark
with which to compare possible “false positive” results. The diagnoses
included multi-infarct dementia, present for at least the previous 7
years, in an 84 year-old diabetic woman. This woman believed that she
was still living with her brother, who died years earlier. She was unable
to take steps of more than two or three inches, even when supported un-
der both shoulders, and was totally dependent on her daughter for feed-
ing, transport, insulin injections and general daily care. She scored
below the cut-offs on the CARB, the WMT and the TOMM but above
the chance range of performance. Another was a woman with Korsakoff’s
syndrome who needed guardianship and trusteeship. A third was a man
with a brain tumor the size of an orange, which had just been resected,
who had recently received 20 whole-brain radiation treatments and who
also had a left temporal lobe hemorrhage. None of these probable false
positive cases would be able to drive safely, all were under 24 hour a
day care from relatives or institutional staff, all had a clear-cut diagnosis
implying extremely severe impairment, and all had clear abnormalities
on CT or MRI brain scans, consistent with their diagnosis. If a person
resembles the latter cases in these respects (e.g., needs constant care, has a definite brain disease, has an abnormal brain scan, and is not allowed to drive), the possibility of false positives on the WMT should be considered. When judging whether a certain patient (e.g., one with a mild head injury and no loss of consciousness) failed WMT despite making an effort, such cases should be borne in mind.

It is equally important to consider some of the cases who are severely impaired but who pass the WMT. For example, Dr. Arthur Shores of Australia (personal communication) reported on a woman with an extremely severe brain injury, whose intelligence score was 56 and whose daily behavior was very degenerated, including smearing feces. This woman easily passed both the WMT and the CARB. Similarly, Dr. Lloyd Flaro (personal communication) reported a boy aged nine years with an IQ of 65, who easily passed CARB and the WMT effort measures. Numerous cases with known severe brain disease who had no difficulty passing the WMT effort measures are described in the manual supplement by Green and Allen (1999). It may be argued that, just as there are now formal studies of many patients showing that failing the WMT indicates incomplete effort, we would also need formal studies to support any claim that people in a given category are unable to pass the WMT despite good effort.

**DAUBERT AND THE WMT**
**IN FORENSIC NEUROPSYCHOLOGY**

Is the WMT admissible under *Daubert*? Clearly, the answer is yes. The WMT is a powerful component in the armamentarium of neuropsychologists seeking to provide scientific evidence for use by the Courts. There are several potential questions listed below, which we can anticipate in *Daubert* challenges to the admissibility of opinions based on the WMT:

*Is the WMT grounded in scientific method?* The critical element of grounding in scientific methodology is empirical verification. From the outset, the WMT has been developed from an empirical scientific perspective. No presuppositions were made that any group would put forth less effort than others. In a lengthy series of studies of differing diagnostic groups, the data have demonstrated the remarkable relationship between WMT effort and neuropsychological test scores.

*Is the procedure reliable and valid?* The WMT has been subjected to intensive scientific validation and cross validation with large samples
drawn from different locations and multiple diagnostic categories. Unlike most research on tests used to address questions of malingering and test validity in forensic settings, the validation of the WMT has been based primarily on actual claimants in forensic settings rather than students and other simulators pretending various conditions.

- Patients with mild head injuries scored lower on the WMT effort tests than patients with severe brain injuries, a finding that can only be explained by lower effort on the part of those with mild head injuries. This is the closest we have to a gold standard in effort testing.
- The cut-offs described in the WMT manual identified with a high level of sensitivity people known to be exaggerating (simulators), but normal controls and patients with moderate to severe brain injuries or neurological diseases easily passed the WMT, demonstrating high specificity for assessing effort.
- Scores on the WMT effort measures were no different in neurological patients with and without verbal memory impairment.
- In 969 patients tested with the computerized WMT, correlations between the three main effort measures (IR, DR and Consistency), ranged from 0.86 to 0.89, showing high internal reliability. The correlations between the WMT memory measures (MC, PA, DFR and LDFR) ranged from 0.71 to 0.90.
- The power of the WMT to predict performance on neuropsychological testing has been demonstrated for a wide variety of tests evaluating different cognitive domains (Table 5). The WMT effort measures explained nearly 50% of the variance in 30,736 test scores (a mean of 34 neuropsychological test scores from each of 904 patients), whereas education explained only 11% of the variance in the same data.
- Alternative explanations have been considered and ruled out in more depth than has been the case with other measures. For example, neither depression, fibromyalgia, chronic pain, neurological diseases nor brain injury were able to explain WMT failure. Failure of the WMT in fibromyalgia patients was associated strongly with financial incentives to exaggerate as part of medical disability claims (35% failure rate), and it was almost non-existent in those with no financial incentive (4% failures).
- The validity of the WMT has been established through comparisons with other symptom validity measures as well as group comparisons and evaluation of individual results. The WMT attained
96% sensitivity relative to the TOMM in a combined sample of 694 patients drawn from three independent sites. The verbal memory format of the WMT and its high face validity may explain why the WMT failure rate in this sample was 2.5 times greater than the TOMM, which utilizes a visual-spatial task.

*What is the error rate?* The WMT enjoys the distinct advantage that it measures a pervasive construct relevant to all neuropsychological tests—effort. The sensitivity, specificity and overall hit rate of the WMT have been demonstrated at impressive levels in several ways. The WMT has been cross validated against other tests associated with malingering or exaggeration. It has been studied by examining and comparing groups with and without incentives to put forth poor effort. It has been applied to patients with objectively demonstrated brain injuries and other neurological conditions in order to verify empirically the ease with which truly impaired individuals can perform the task. As described above, there is abundant evidence of the accuracy of the WMT in detecting poor effort.

*Has the procedure been subjected to peer review?* Because the WMT is a relatively new test, some of the validation studies are still under review, but some studies have already been published (e.g., Gervais, Green, Allen, & Iverson, 2001; Gervais, Russell, Green, Allen, Ferrari, & Pieschl, 2001; Green, 2001; Green & Iverson, 2001; Green, Iverson, & Allen, 1999; Green, Rohling, Lees-Haley, & Allen, 2001; Iverson, Green, & Gervais, 1999), and a number of other papers have recently passed the journal peer review process and are currently in press. Dozens of studies have been submitted as poster presentations and abstracted in *Archives of Clinical Neuropsychology*.

*Is the procedure generally accepted in the scientific community?* The claim that assessment of effort is relevant and helpful to the trier of fact in forensic neuropsychological cases is beyond impeachment. The proposition that examinees must do their best on tests has been well recognized and generally accepted in the psychological testing community for the entire history of psychological testing (e.g., see Anastasi, 1954; Cronbach, 1949; Terman & Merrill, 1937). The overwhelming majority of psychological and neuropsychological assessment reports comment on the level of effort exerted by the examinee. The special value of the WMT is to add an empirical test of effort to the examiner’s clinical impressions. The basic principles incorporated into the WMT have enjoyed wide acceptance in the neuropsychological community for years (e.g., Binder, 1993; Pankratz, Fausti, & Peed, 1975). The use of the
WMT itself is growing rapidly in many countries, both in English and in translations.

**Falsification:** *Can the expert’s opinion, conclusion, or theory be tested?* The thesis that the WMT measures effort can be and has been tested repeatedly. For example, this assertion has been studied through comparison of levels of performance of WMT pass/fail groups on other measures of neuropsychological functioning. It has been assessed by comparison with other measures associated with test validity. It has been studied by examining the plausibility of scores, e.g., the finding that severe brain injury patients and other objectively impaired individuals can pass the WMT whereas financially motivated mild injury claimants often perform poorly.

The above arguments apply mainly to the validity of the WMT effort subtests as measures of effort. The use of the WMT to measure memory and the validity of the memory subtests is a separate and very large topic. It may be seen from Tables 1 and 2 that the WMT memory subtests are sensitive to the effects of brain injury and to neurological diseases, which cause impaired verbal memory. However, this paper is not meant to address the reliability and validity of the WMT memory subtests (MC, PA, DFR and LDFR), which will be the subject of another article.

**CONCLUSIONS**

Hartman (in press) states in a review of the WMT, “Based on current and upcoming publications, it appears that the *Word Memory Test* and its cousin, the CARB, published by the same company, will soon be the most researched effort tests available. Considering the preliminary findings of such research, it seems likely that future investigations of litigating populations, (e.g., fibromyalgia, multiple chemical sensitivity, mild traumatic brain injury) which omit such tests may come to be considered unpublishable and uninterpretable.” Unlike other effort test scores, WMT scores have been shown to be major predictors of scores on many neuropsychological tests in large clinical samples. It is this aspect of the WMT research that is its major strength. Because of the WMT research, we now know that effort has a much larger effect on test scores in general than severe brain injury, at least in cases involved in compensation claims. Effort explains so much of the variance in other test results that the validity of neuropsychological results might be questioned in any case presented in a court of law, if measures of effort
are not used to rule out invalid test data. Effort tests at present may be thought of as analogous to quality control instruments—a sort of Daubert challenge to the plaintiff that throws up a red flag saying, “Data from this plaintiff cannot be trusted.” If a plaintiff produces invalid results on one or a few tests, it renders the other data doubtful as evidence of the patient’s true abilities. Similarly, studies of various patient groups show that any group differences in test scores obtained without controlling for effort may be questioned, especially if some of the patients are engaged in a financial compensation claim.

REFERENCES


