Green, P., Berendt, J., Mandel, A., & Laane, L. (10/1/03). Comparison of the Test of Memory Malingering and the Word Memory Test for identifying response bias in a series of compensation cases. *WebPsychEmpiricist*. Retrieved (date) from: http:/wpe.info/papers_table.

WPE WebPsychEmpiricist

Comparison of the Test of Memory Malingering and the Word Memory Test for identifying response bias in a series of compensation cases.

10/1/03

Paul Green,

Neurobehavioral Associates, Edmonton, Alberta, Canada

John Berendt, Allen Mandel,

Mandel and Associates, Calgary, Alberta, Canada

Lyle Allen,

Durham N.C., U.S.A.

Abstract

Most symptom validity tests (SVT) are objectively very easy. However, two tests which are equally easy will not necessarily be equally sensitive to cognitive exaggeration. In the present study, we gave a series of compensation claimants two SVTs, one based on forced choice recognition of visual stimuli, the Test of Memory Malingering (TOMM; Tombaugh, 1996) and the other based on word recognition memory, the Word Memory Test (WMT; Green, Allen & Astner, 1996). Forty patients (28% of all cases) failed the TOMM, the WMT or both. The overall agreement rate was 82% between the TOMM and the WMT, with regard to valid or invalid effort, leaving disagreement in 18% of all cases. There was only one case (0.7%) who passed the WMT and failed the TOMM. However, there were 25 cases (17.4%) who passed the TOMM but failed the WMT. The verbal memory test scores of the latter cases were found to be significantly lower than the scores from 112 patients with moderate-severe brain injuries or neurological diseases. It was concluded that the WMT was more sensitive to suboptimal effort than the TOMM. These results replicate findings from a previous independent study.

Correspondence should be addressed to Dr. Paul Green, 201, 17107-107 Ave., Edmonton, Alberta, Canada, T5S 1G3 email: paulgreen@shaw.ca

Comparison of the Test of Memory Malingering and the Word Memory Test for identifying response bias in a series of compensation cases.

10/1/03

The use of symptom validity tests (SVT) in neuropsychological assessment has become indispensable, in the wake of numerous studies showing a substantial amount of exaggeration of cognitive impairment in people being given neuropsychological tests (Larrabee, 2000). In a review of tests for detecting exaggerated cognitive deficits, Iverson & Binder (2000) stated that it is essential to employ methods for ruling out invalid test results in any compensation related neuropsychological assessment. Cognitive testing is applied by psychologists to people with many different diagnoses and there is no a priori reason why the concept of measuring effort to determine whether or not test data are valid should be limited to any particular diagnostic group. In a study of 904 consecutive outpatients with diagnoses as varied as mild head injury, severe brain injury, stroke, brain tumor, major depression, fibromyalgia, pain disorder and orthopedic injuries, it was shown that effort was inadequate to produce valid test results in approximately 30% of all cases (Green, Rohling, Lees-Haley & Allen, 2001). Effort explained 50% of the variance in the whole battery of tests, representing an average of 34 test scores per patient. Effort had a far greater impact on neuropsychological test scores than severe brain injury or brain tumors, leading to the conclusion that measuring effort is essential whenever neuropsychological tests are administered. The major effect of low effort in suppressing a wide range of test scores occurred equally in all diagnostic groups.

Several other studies have identified exaggeration of cognitive difficulties in a substantial proportion of disability claimants presenting with chronic pain (Allen, Conder, Green, & Cox, 1997; Kay & Morris-Jones, 1998; Schmand, et al., 1998; Schnurr & MacDonald, 1995) and somatoform symptomatology (Boone & Lu, 1999). Gervais, Russell, et al., (2001) studied two groups of patients with fibromyalgia, one working and one claiming disability. The rate of failure on effort tests was only 4% in those who were working but 35% of the disability claimant group failed one or both effort tests. These results suggested that symptom exaggeration was a function of motivation and not diagnostic group. Gervais, Green, Allen, & Iverson (2001) argued that baseline failure rates in excess of 40% on each of two separate SVTs in patients with DSM-IV pain disorder were not a function of pain or emotional distress but a sign of symptom exaggeration or poor motivation. This conclusion was reached because coaching of subsequent pain patients led to a greatly reduced failure rate on the SVT about which they had been warned, while the failure rate was unchanged on a second, equally easy SVT, about which no warning was given.

Many methods for detecting exaggeration of cognitive deficits are reviewed by Iverson and Binder (2000) and Sweet (1999), including the use of algorithms derived from scores on tests of memory or problem solving. There has been a proliferation of forced-choice tests designed to detect exaggerated impairment, including several digit recognition tests (Allen, Conder, et al., 1997; Binder & Willis, 1991; Conder, Allen & Cox, 1992; Green & Iverson, 2001; Hiscock & Hiscock, 1989; Slick, Hopp, Strauss, Hunter, & Pinch, 1994; Slick, Hopp, Strauss & Spellacy, 1996). The Test of Memory Malingering (TOMM; Tombaugh, 1996) is a forced choice *visual spatial recognition* memory test and it was used in the current study. Others, such as the 21 Item Test described by Iverson, Franzen and McCracken (1991, 1994) employ forced choice verbal recognition memory tasks.

The need for a test which would be relatively resistant to prior coaching and which would simultaneously provide measures of verbal memory and effort gave rise to the Word Memory Test (WMT; Green, Allen & Astner, 1996; Green & Astner, 1995). The WMT effort subtests rely upon forced-choice recognition of words previously presented in pairs. Each of the first four subtests also serves, in part, as a learning trial prior to free recall of the list, making it possible to test effort and various aspects of memory simultaneously. Support for the WMT being relatively resistant to the effects of prior coaching came from the inability of sophisticated simulators to produce believable patterns of deficits on the test, despite being told that it was designed to detect exaggerated memory deficits (Iverson, Green & Gervais, 1999) and also from a study of coaching in patients with Pain Disorder (Gervais, Green, et al., 2001). Numerous validation studies of the WMT are reviewed in Green, Lees-Haley and Allen (2002).

Most SVTs are similar to each other, in that people with severe traumatic brain injury and other serious and objectively verifiable neurological impairment can easily pass these tests if they make an effort to do so. The majority of forced-choice SVTs have strong ceiling effects, which serves to minimize the rate of false positive identification of poor effort. Studies of patients with severe brain injuries and neurological diseases have consistently led to mean scores well above the cut-offs for incomplete effort on each of the three effort measures in the WMT

(Allen & Green, 1999; Green, Iverson & Allen, 1999; Iverson, et al., 1999). There were no differences on the WMT effort measures between neurological patients with and without impaired verbal memory (Green & Allen, 1999). The cutoffs for poor effort on the WMT were set at approximately three standard deviations below the combined mean scores from patients with severe traumatic brain injury (Allen & Green, 1999) and neurological diseases, such as brain tumors and ruptured aneurysms (Green & Allen, 1999). It has been shown that, in children tested clinically, the mean scores on the WMT effort measures were no different from those of a group of adults who were motivated to gain custody of their children (Green & Flaro, 2003). The WMT cutoffs are listed in the test manuals but not reported here for test security purposes. Similarly, using the TOMM, Tombaugh (1996) showed that 121 patients with known cognitive impairment could easily score above the trial two cut-off for inadequate effort. Therefore, except in extreme cases of impairment, such as advanced dementia, if a person fails either the TOMM or the WMT, doubt is cast upon the validity of the person's other test results.

We cannot assume that any two effort tests are of equivalent sensitivity. Two tests may both be relatively insensitive to genuine neurological impairment but this does not mean that they will necessarily be equally sensitive to the exaggeration of cognitive deficits. In order to study their differential sensitivity to response bias, we have to study the results of more than one effort test in the same subjects. One aspect of face validity of SVTs is the extent to which patients perceive the test to be relevant to their claimed disability. Gervais, Green & Allen (1999) hypothesized that people claiming prominent visual memory difficulties would be most likely to fail a visual SVT (TOMM), whereas those emphasizing numerical impairment would more often fail a digit recognition test (CARB) and those reporting mainly verbal memory impairment would selectively fail a verbal recognition memory test (WMT). They used a standardized self-report measure to assess memory complaints in 150 patients with various diagnoses involved in compensation claims, including a subgroup of patients with DSM-IV pain disorder. Contrary to expectation, irrespective of patients' self-reported memory problems, a substantial majority failed the verbal test (the WMT), compared with the visual test (the TOMM). The failure rate on the digit recognition test (CARB) was intermediate between those of the WMT and the TOMM. A similar excess of verbal versus visual-spatial memory complaints on the Memory Complaints Inventory was found in patients with mild head injuries or severe traumatic brain injuries (Green, Allen & Iverson, 1999). This pattern was seen both in those who passed and in those who failed the effort measures. Therefore, we might hypothesize that people who exaggerate their memory impairment will be more likely to fail verbal SVTs, which might be perceived as being most relevant to their clinical complaints, while less often failing an easy visual-spatial test, such as the TOMM, if it is perceived as less relevant to their memory complaints.

The purpose of the present study was to test the latter hypothesis by comparing failure rates on the WMT and the TOMM, when both tests were given to a consecutive series of patients assessed as part of a compensation-related disability claim. Using a sample of patients from a different city, tested by different examiners, we attempted to replicate results previously reported by Gervais et al. (1999), who found significantly different failure rates on the TOMM and the WMT. Before becoming aware of the results of Gervais et al. (1999), the second and third authors independently administered both the WMT and the TOMM to 144 people being evaluated for eligibility for disability benefits. To strengthen the analysis and to put the patients' memory test performances into perspective, the results from these patients were compared with those of a comparison group of 112 patients with known cognitive impairment from moderate to severe traumatic brain injuries or neurological diseases (the TBI-NEURO group), who were assumed to be making a reasonable effort to perform well for reasons given below.

Method

Participants: Two groups of patients were involved in this study and they will be called the "Disability" group (*n* = 144) and the traumatic brain injury or neurological disease group ("TBI-NEURO", *n* = 112). The Disability claimant group consisted of 144 consecutive patients, who claimed to have a disability affecting their ability to work, and the purpose of the assessments was to evaluate psychological disability and the potential for alternative employment after an illness or injury. Forty-five patients (31.3%) were involved in litigation, 93 cases (64.6%) were candidates for a Federal Government medical disability program and 6 cases (4.2%) were seeking worker's compensation or insurance disability benefits. Forty-one (28.5%) of the Disability subjects had orthopaedic injuries, 36 cases (25%) were diagnosed with fibromyalgia or chronic pain syndrome, 37 (25.7%) with multiple medical difficulties, 15 (9.1%) with a head injury or neurological problem and eight (5.6%) reported primarily emotional difficulties. Two cases (1.4%) were diagnosed with chronic fatigue syndrome and seven (4.9%) were ex-residents of an institution for the mentally retarded, whose intelligence varied from mentally handicapped to low average. The mean measured WAIS-R full scale IQ in the Disability group was 91 (*SD* = 14.8) and the mean age was 39.3 (*SD* = 10.9). The Disability patients had a mean of 10.8 years of education (*SD* = 2.4) and 43% were men. Spearman rank order correlations were computed between the demographic data and the four SVT measures used in the study, which are not normally distributed. As reported by previous investigators, there were no significant correlations between any of the effort test measures and age ($p > 0.52$ for all), gender ($p > 0.14$ for all), years of education ($p > 0.29$), handedness ($p > 0.48$) or incidence of English as a second language ($p > 0.21$). All of the patients in this study were fluent in English and none needed an interpreter during the assessment.

The TBI-NEURO control group consisted of 67 moderate to severe brain injury patients and 45 patients with neurological diseases, including brain tumors, ruptured aneurysms, strokes and multiple sclerosis. The TBI-NEURO patients had all been given the WMT but not the TOMM as part of a comprehensive clinical neuropsychological assessment related to workers' compensation, medical disability or personal injury claims in the private clinical neuropsychology practice of the first author. Fifty-seven of these patients with severe traumatic brain injury (85%) were described in detail by Allen & Green (1999), and 40 of the neurological patients (89%) were described by Green & Allen (1999). In the latter manual supplements, the goal was not to study how patients performed when making a poor effort. On the contrary, it was to obtain normative data on the WMT from patients with moderate to severe brain injuries or neurological diseases of the brain, who were assumed to be making a valid effort. Therefore, when there was independent data from other sources strongly indicating incomplete effort (e.g. a worse than chance score on a test such as Warrington's Face Recognition Memory Test), the cases were excluded. The reasons for exclusion were described for each single case in the manual supplements and none were excluded based on their CARB or WMT scores (Allen & Green, 1999; Green & Allen, 1999). Ten new cases of moderate to severe traumatic brain injury and five new neurological cases were added to the latter groups, to form the current TBI-NEURO group (*N* = 112). This group consisted of 112 out of 122 cases after excluding a total of only 10 patients (i.e., 8% of all cases). The 112 TBI-NEURO patients were assumed to be making a full effort, consistent with their almost perfect scores on the WMT effort measures.

The TBI-NEURO group was 71% male, had an average age of 42.6 years (*SD* = 11.3), and an average education of 11.8 years (*SD* = 3.9).

Procedure: All Disability cases in the current study were given both the WMT and the TOMM as part of their clinical evaluation. The TOMM involves the presentation of 50 line drawings of objects in a booklet, followed by a recognition trial in which the person is required to choose the original picture from a pair of drawings. After a delay, the person is shown the items again and a second test trial is given. In the test manual, Tombaugh (1996) states that patients with known cognitive impairment can usually score at or above a certain level out of 50 correct on trial two, such that a score below this level is regarded as a sign of questionable effort, and this was the cut-off for poor effort used in the present study (cutoff not specified here for test security purposes). Worse than chance scores on any TOMM subtest were also used to define failure but all such cases scored less than the standard cut-off on trial 2.

In the computerized Word Memory Test (WMT), the subject is presented with 20 semantically linked simple word pairs (e.g., dog - cat), each pair appearing on the screen for six seconds. After the list is presented twice, there is an immediate recognition trial (IR), in which the person is shown new word pairs containing only one of the words from the original list and must select the words belonging to the original list (e.g., "dog" from the pair "dog - rabbit"). Without advance warning, a similar delayed recognition procedure (DR) is administered after 30 minutes, using different foil words paired with the original words (e.g., "dog" – "rat"). WMT scoring includes a calculation of consistency of responding from IR to DR (called "Consistency" in the tables below) and this measure is also sensitive to patient effort. Incomplete effort was concluded based the cutoffs for IR and DR recommended by the test authors in the latest supplement for the WMT manual (Allen & Green, 1999; Green and Allen, 1999). It was also based on the consistency between scores on these two subtests.

The WMT delayed recognition trial is followed immediately by a series of tests of gradually increasing difficulty, which measure *verbal memory*. These include multiple choice (MC), in which the person is shown the first word of each pair and has to select the other word from eight options; paired associates (PA), in which the tester speaks the first word and asks for the second word in each pair; delayed free recall (DFR), involving the person saying as many words as possible from the list in any order, while the tester records responses on the computer, and long delayed free recall (LDFR), which is the same task as DFR but follows another twenty minute delay. Visual and/or auditory feedback is given during the MC and PA subtests. All raw WMT subtest scores and effort calculations are expressed as a percent of the maximum possible score. The examiner was present throughout both the WMT and TOMM administrations and so any differences in failure rates between the TOMM versus the WMT could not be explained as being due to problems understanding the task or observer effects.

Results

The rates of failure on the TOMM versus the WMT were compared in the main clinical sample of 144 cases in the Disability group and are shown in Table 1. There were 104 cases (72%) who passed both TOMM and all three WMT effort measures. The overall agreement rate was 81.9% between the TOMM and the WMT, with regard to valid or invalid effort. There were 15 (10.4% of all cases) who failed TOMM and 39 (27% of all cases) who failed the WMT. There were 25 (17.4% of all cases) who passed the TOMM and failed the WMT. If randomly inconsistent responding were largely responsible for the observed difference, we would expect to find a large number of patients failing the TOMM and passing the WMT. However, there was only one case (0.7%) who passed the WMT and failed the TOMM. No further analysis was undertaken using this single patient's data.

A central question raised by these results was whether, in the 25 cases who passed the TOMM and failed the WMT, the effort being made was sufficient to produce valid results? Were these cases false positive classifications of poor effort by the WMT or false negatives for TOMM? A third possibility, which is not necessarily inconsistent with the first two hypotheses, might be that these 25 patients were inconsistent in their effort, making a good effort on TOMM but a poor effort on WMT. However, effort tests are used mainly to decide whether other test results are likely to be valid or not and so the clinically important question is what effort did these patients make on ability tests, as opposed to the effort measures represented by the TOMM and the WMT? Also, in cases who were inconsistent because they passed the TOMM but failed an equally easy test, the WMT, were other test results plausible or were they invalid as indicators of their true abilities? To shed light on this question, further comparisons were undertaken using the TBI-NEURO group as a benchmark.

Table 2 shows the mean scores on each of the four WMT memory subtests and three effort measures from four groups, which are: the Disability patients who passed both the TOMM and the WMT; the TBI-NEURO group; the Disability patients who failed the WMT and passed the TOMM; and the Disability patients who failed both the TOMM and the WMT. No results are shown from the group which failed the TOMM and passed the WMT because there was only one case in this group. It may be noted that, for all seven WMT subtests, the rank order of scores across groups is identical, with the highest scores in Disability patients passing both the TOMM and the WMT and the lowest scores in those failing both of these tests. The most important finding in Table 2 is that, whether or not they failed the TOMM, those who failed the WMT effort measures scored lower than the TBI-NEURO group on all four memory subtests (multiple choice, paired associates, delayed free recall and long delayed free recall).

To compare these four groups with each other statistically, we averaged the WMT effort measures to create a single effort score (Table 3). Also, the WMT memory subtest scores were averaged to create a mean verbal memory score. Because scores on the WMT effort measures are not normally distributed, all between group comparisons of performance on the dependent measures were performed with non-parametric statistics using the Mann-Whitney Test. Comparing each of the four groups with one another on the average WMT effort and memory measures required 12 separate two-tailed comparisons. A Bonferroni correction (0.05 / 12 = 0.0042) was adopted for multiple comparisons to reduce the chance of identifying spurious differences ($p <= 0.004$).

<u>Effort measures:</u> The TBI-NEURO group produced significantly higher WMT mean effort scores than both (a) Disability patients failing the WMT and TOMM ($U = 17.0$; $z = -5.98$; $p < 0.0005$) and (b) Disability patients failing only the WMT ($U = 78$; $z = -7.38$; $p < 0.0005$). In contrast, Disability patients passing the TOMM and the WMT showed significantly higher WMT effort scores than the patients failing both the WMT and TOMM ($U = 0$; $z = -6.13$; $p < 0.0005$) and patients failing only the WMT ($U = 4.5$; $z = -7.78$; $p < 0.0005$). The TBI-NEURO and the Disability patients passing the WMT were not significantly different from one another on the mean WMT effort score ($U = 4602.5$; $z = -2.49$; $p < 0.02$). Patients who failed both the TOMM and the WMT had a significantly lower average WMT effort score than patients who failed only the WMT ($U = 73.5$; $z = -2.9802$; $p < 0.003$).

Memory subtest scores: On the average WMT memory score, 5 of the 6 between group comparisons were highly significant ($p < 0.0005$). The only groups that did not differ from each other in their performance on the WMT memory measures were the Disability patients who failed both the TOMM and the WMT and the Disability patients who failed the WMT, while passing the TOMM ($U = 114$; $z = -1.79$; $p > 0.07$). Thus, failing the WMT effort measures was associated with equally low memory scores, whether or not the TOMM was failed.

Disability patients passing the TOMM and WMT scored significantly higher than the TBI-NEURO group on the memory scores, as expected ($U = 3380.5$; $z = -4.32$; $p < 0.0001$). However, the TBI-NEURO group produced significantly higher mean memory scores than the Disability patients who failed the WMT and TOMM ($U = 96.5$; $z = -5.22$; $p < 0.0001$). Their scores were also higher than those of patients failing only the WMT ($U = 371$; $z = -5.45$; $p < 0.0001$). Disability patients, who passed both the TOMM and the WMT produced significantly higher WMT memory scores than those failing the WMT and TOMM ($U = 26.50$; $z = -5.83$; $p < 0.0001$). They also scored higher than patients failing only the WMT ($U = 162.5$; $z = -6.76$; $p < 0.0001$).

TOMM scores and correlations: In all Disability cases, the mean TOMM trial one score was 45.8, *SD* 6.25 and the mean trial two score was 48.2, *SD* 4.8. The mean scores on TOMM trials one and two in the Disability patients were as follows: (a) cases passing TOMM and WMT (n = 104), trial one mean = 48.1 *SD* 2.8, trial two mean = 49.8, *SD* 0.7; (b) cases failing both ($n = 14$), trial one mean = 30.0, *SD* 5.7, trial two mean = 34.9, *SD* 5.7; (c) cases passing TOMM and failing WMT (n = 25), trial one mean = 45.0, *SD* 3.3, trial two mean = 49.0, *SD* 1.4 and (d) the single case failing TOMM and passing WMT, trial one = 40, trial two = 42.

The correlation between the mean score on TOMM trials one and two was 0.9 and the correlation between WMT IR and DR was 0.86. TOMM scores were highly intercorrelated with both WMT effort and memory scores. TOMM trials one and two correlated with the mean WMT effort measure at 0.78 and 0.76 respectively (both $p < .0005$). The mean of TOMM trials one and two correlated with the mean memory score on the WMT at 0.6 ($p < .0005$). For comparison, the mean of the three WMT effort measures correlated with the mean WMT memory score at 0.77.

<u>Intelligence, age and education:</u> It has previously been reported that there is no significant relationship between WMT effort scores and age, education or intelligence, which is consistent with the proposition that they measure effort and are minimally sensitive to variations in ability (Allen, Conder, et al., 1997; Green, et al., 1996;). ANOVA analysis of demographic data revealed that there were no differences between the two groups in age ($p > 0.10$), years of education ($p > 0.14$), handedness ($p > 0.14$) or incidence of English as a first language ($p > 0.14$). 44% of the Disability patients and 71.4% of the TBI-NEURO group were men, which is a significant difference. However, Spearman rank order correlations which included all subjects showed that gender was uncorrelated with performance on any of the WMT or TOMM effort measures ($p > 0.14$ for all).

Because some of the Disability subjects were of low intelligence, the effect of intelligence on each of the SVT scores was examined in all 91 cases for whom IQ scores were available. The group was divided into those with a FSIQ less than 85 (mean 75.5, *SD* 7.4, n = 33) and those with a FSIQ of 85 or higher (mean 97.1, *SD* 9.0, n = 59), a mean difference of 21.6 full scale IQ points. There were no significant differences between these groups in their mean scores on the TOMM trial two measure used in this study (mean 46.5, *SD* 6.4 in higher IQ cases versus a mean of 48.1, *SD* 4.2 in lower IQ cases; $F = 1.6$, *df* 1,90, $p = 0.2$). Similarly, there were no differences between these groups on the WMT immediate recognition trial (mean 87.2, *SD* 14.7 in higher IQ cases versus a mean of 86.1, *SD* 14.4 in lower IQ cases; $F = 0.12$, *df* 1, 90, $p = 0.7$), no differences on the WMT delayed recognition trial (mean 87.1, *SD* 17.6 in higher IQ cases versus a mean of 86.4, *SD* 14.9 in lower IQ cases; $F = .04$, *df* 1, 90, $p = 0.8$) and no differences on WMT consistency (mean 84.3, *SD* 16.8 in higher IQ cases versus a mean of 86.4, *SD* 14.9 in lower IQ cases; $F = 0.5$, $p = 0.48$).

Discussion

In this series of patients with heterogeneous diagnoses, there was overall agreement between the TOMM and the WMT in 81.9% of cases about whether or not patients were making a sufficient effort to produce valid results. This finding supports the conclusion that the TOMM and the WMT effort measures are assessing a similar underlying construct, which is effort rather than ability. In keeping with this conclusion, there was no significant difference on any of the

WMT effort measures or on the TOMM between Disability patients with a mean full scale IQ of 75 and those with a mean full scale IQ of 97. Nor was there any difference between the WMT effort scores in the brain injured and neurological patients (TBI-NEURO), compared with the Disability patients who passed the TOMM and the WMT (Table 3). The fact that the brain injury and neurological cases obtained mean recognition scores above 95% correct on the WMT recognition effort measures shows that these measures are largely insensitive to cognitive impairment arising from brain injury or disease (Table 2). Hence, failure on the WMT effort measures by some of the Disability patients is likely to reflect response bias or insufficient effort to produce valid results.

Supporting the latter conclusion, the Disability patients who failed either the TOMM or the WMT produced scores on the verbal memory measures of the WMT which were implausible and invalid. This is shown by the fact that their memory scores were significantly worse than those of a group of 112 patients with either moderate to severe traumatic brain injuries or neurological diseases, including brain tumors, ruptured aneurysms, strokes and multiple sclerosis (Table 3). Yet, as expected, the TBI-NEURO patients produced impaired verbal memory scores compared with 104 Disability patients, who passed both TOMM and the WMT.

In the Disability patients, 72% of cases passed both SVTs . Those who passed the WMT effort measures also passed the TOMM, except for one case. However, there were 25 patients in which the results of the TOMM and the WMT were discrepant because they passed the TOMM, suggesting satisfactory effort, but failed the WMT effort subtests, suggesting the opposite. The results of these patients on verbal memory measures were extremely low as a result of suboptimal effort. As shown in Table 3, these Disability patients scored significantly lower than the TBI-NEURO group on the WMT memory subtests, which measure ability. Based on their diagnostic categories outlined above, it is thought to be extremely unlikely that these Disability patients actually had significantly worse verbal memory than the TBI-NEURO group. The most likely explanation for the low verbal memory test scores observed in these patients would be incomplete effort or exaggeration of cognitive difficulties.

The results lead us to the conclusion that a substantial number of patients producing invalid test results in a clinical setting will pass the TOMM but fail the WMT effort measures and that, in such cases, other test results are likely to be of doubtful validity. It may be noted that the WMT effort measures and the TOMM are of approximately equal difficulty, with scores between 95% and 100% correct in Disability cases passing both tests. Both the TOMM and the WMT effort measures have been found to lead to very high scores in patients with known brain damage or disease. Therefore, passing one test but failing the other is, in itself, evidence of inconsistency of performance, suggesting unreliable and invalid results. This adds further support to the conclusion that those who passed the TOMM but failed the WMT were putting forth insufficient effort to produce valid results.

When failure on either SVT is taken as evidence of poor effort, the use of the TOMM alone failed to identify 62.5% of the patients demonstrating poor effort. The present results are very similar to those from an independent patient sample studied by Gervais et al. (1999), in whom there was a 32% rate of failure on one or more of the three SVTs used. These authors found that 70% of 48 patients who failed either CARB or the WMT passed the TOMM, while not a single case failed the TOMM and passed the WMT. From these two studies, it seems that failing TOMM and passing the WMT is a rare event.

The 28% base rate of SVT failure in the current sample is similar to those reported in other samples. For example, there was a 30% failure rate for CARB alone in 1752 protocols drawn from 13 sites in the US and Canada (Allen, Richards, Green, Iverson & Conder, 1997). The relative insensitivity of the TOMM compared to the WMT might be reduced in samples characterized by much higher rates of SVT failure. This hypothesis is supported by the fact that patients who failed the TOMM and WMT showed significantly greater exaggeration on the WMT effort measures than patients who only failed the WMT. It is possible that less sophisticated or more poorly motivated patients will make less of a distinction between these two symptom validity tests.

In this study, poor effort on the WMT resulted in lower scores on verbal memory subtests than found in patients with severe brain injuries or neurological diseases but no visual memory test was used. It could be argued that some exaggerators might choose to fail visual memory tests and that, in such cases, the TOMM would have been more sensitive than the WMT to poor effort. However, this is contrary to the data because, with one exception, all cases who failed the TOMM also failed the WMT and there were no such cases in the study of Gervais et al. (1999). Therefore, no matter what external ability measure had been used, it would have been impossible for the TOMM to have been more sensitive to exaggeration than the WMT.

A major excess of failures on the WMT effort subtests versus the TOMM has now been found in two independent studies of patients assessed as part of compensation or disability claims. This suggests the need for further research into the reasons for the apparently greater sensitivity of the WMT over the TOMM in detecting poor effort. It is possible that verbal effort tests are more sensitive to cognitive exaggeration than visual SVTs owing to patient phenomenology. Previous studies have shown that complaints of impaired verbal memory are far more frequent than visual-spatial memory complaints (e.g., Gervais, et al., 1999; Green & Allen, 2000). This might help to explain why people who are exaggerating their cognitive difficulties are more likely to fail a verbal effort test, such as the WMT, than a visual effort test, such as the TOMM. Perceived difficulty is another variable that could affect performance on SVTs which utilize different types of stimuli and it is possible that the WMT seems more difficult than the TOMM. In the current study, we did not examine self-reported memory problems or the perceived difficulty of the tests and so the results do not shed light on these questions.

The implications of a major difference in failure rates on one type of effort test versus another are important for clinical practice. Many neuropsychologists use only one effort test. Our data show that it cannot be assumed that two examiners will arrive at an acceptable level of agreement about the incidence of incomplete effort in the same patients if they use different SVTs. The choice of one test rather than another could make a large difference to the proportion of cases estimated to be producing invalid test results. We found that the WMT failure rate was 160% greater than the failure rate on TOMM (see Table 1). The failure rate on effort tests is influenced not only by the number of people in the sample who are prone to exaggerating cognitive deficits (true base rate), but also by the type of effort tests which are administered and their sensitivity to exaggeration. Because it is important to establish uniformity in the classification of poor effort by different examiners, more research is needed to evaluate the rates of agreement between the many different methods that are currently employed for the evaluation of patient effort. This study cannot tell us why people fail different effort tests at different rates but it makes it clear that it would be very unwise to assume that two methods for evaluating effort are equally sensitive to exaggeration without having strong supportive evidence to show that this is the case.

References

Allen, L.M., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB '97 Manual for the Computerized Assessment of Response Bias*. CogniSyst, Inc., Durham, NC.

Allen, L.M., & Green, P. (1999). Severe TBI Sample Performance on CARB and the WMT. In Green, P., & Allen, L.M. (Eds.). *1999 Supplement for the Computerized Assessment of Response Bias, Word Memory Test and Memory Complaints Inventory*. CogniSyst, Inc., Durham, N.C.

Allen, L.M., Richards, P.M., Green, W.P., Iverson, G.L., & Conder, R.L., Jr. (1997). Performance patterns on the Computerized Assessment of Response Bias in 1752 compensation cases (Abstract). *Archives of Clinical Neuropsychology*, *13*(1), 15-16.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment*, *3*, 175-181.

Boone, K.B., & Lu, P.H. (1999). Impact of somatoform symptomatology on credibility of cognitive performance. *The Clinical Neuropsychologist*, *13*(4), 414-419.

Conder, R., Allen, L., & Cox, D. (1992). *Manual for the Computerized Assessment of Response Bias*. CogniSyst, Inc., Durham, N.C.

Gervais, R., Green, P., & Allen, L.M. (1999). Differential sensitivity to symptom exaggeration of verbal, visual, and numerical symptom validity tests (Abstract). *Archives of Clinical Neuropsychology*, *14*(8), 746-747.

Gervais, R., Green, P., Allen, L.M., & Iverson, G. (2001). Effects of coaching on symptom validity testing in chronic pain patients presenting for disability assessments. *Journal of Forensic Neuropsychology*, *2* (2), 1-19.

Gervais, R.O., Russell, A.S., Green, P., Allen, L.M., Ferrari, R., & Pieschl, S D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology*, *28*, 1892-1899.

Green, P., & Allen, L.M., (1999). Performance of Neurological Patients on the Word Memory Test (WMT) and Computerized Assessment of Response Bias (CARB). In Green, P., & Allen, L.M. (Eds.). *1999 Supplement for the Computerized Assessment of Response Bias, Word Memory Test and Memory Complaints Inventory*. CogniSyst, Inc., Durham, N.C.

Green, P., & Allen, L.M., (2000). Patterns of memory complaints in 577 consecutive patients passing or failing symptom validity tests. (Abstract) *Archives of Clinical Neuropsychology*, *15* (8), 844-845

Green, P., Allen, L.M., & Astner, K. (1996). *The Word Memory Test: A User's Guide to the Oral and Computer-Administered Forms, US Version 1.1*. CogniSyst, Inc., Durham, NC.

Green, P., Allen, L.M., & Iverson, G.L. (1999). Utility of the Memory Complaints Inventory for identifying symptom exaggeration in mild to moderate traumatic brain injury (Abstract) Archives *of Clinical Neuropsychology*, *14*(8), 743-744.

Green, P., & Astner, K. (1995). *Manual for the Oral Word Memory Test*. Neurobehavioural Associates, Edmonton, Alberta, Canada.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology, 9* (3), 189-207.

Green, P., & Iverson, G. (2001). Validation of the Computerized Assessment of Response Bias in litigating patients with head injuries. *The Clinical Neuropsychologist, 15* (4), 492-497.

Green, P., Iverson, G. L., & Allen, L.M. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, *13*(10), 813-819.

Green, P., Lees-Haley, P.R., & Allen, L.M. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology*, *2*(3/4), 97-124

Green, P., Rohling, M.L., Lees-Haley, P.R., & Allen L.M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15* (12), 1045-1060.

Hiscock, M., & Hiscock, C.K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, *11*, 967-974.

Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation*, *15*(2), 829-858.

Iverson, G. L., Franzen, M. D., & McCracken, L. M. (1991). Evaluation of a standardized instrument for the detection of malingered memory deficits. *Law and Human Behavior*, *15,* 667-676.

Iverson, G. L, Franzen, M. D., & McCracken, L. M. (1994). Application of a forced choice procedure designed to detect malingered memory deficits. *Archives of Clinical Neuropsychology*, *9,* 437-450.

Iverson, G., Green, P., & Gervais, R. (1999). Using the Word Memory Test to detect biased responding in head injury litigation. *The Journal of Cognitive Rehabilitation, 17*(2), 4-9.

Kay, N. R. M., & Morris-Jones, H. (1998). Pain clinic management of medico-legal litigants. *Injury*, *29*, 305-308.

Larrabee, G. J. (2000). Neuropsychology in Personal Injury Litigation. *Journal of Clinical and Experimental Neuropsychology*, *22* (5), 702-707

Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, H. L. (1998). Cognitive complaints in patients after whiplash injury: The impact of malingering. *Journal of Neurology, Neurosurgery, and Psychiatry*, *64*, 339-343.

Schnurr, R., & MacDonald, M. (1995). Memory complaints in chronic pain. *The Clinical Journal of Pain*, *11*, 101-111.

Slick, D.J., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's forced choice memory test. *Journal of Clinical and Experimental Neuropsychology*, *16*, 472-481.

Slick, D.J., Hopp, G., Strauss, E., & Spellacy, F.J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology*, *18*, 911-922.

Sweet, J.J. (1999). Malingering: differential diagnosis. In Sweet, J.J. (Ed), *Forensic Neuropsychology Fundamentals and Practice*. Lisse: Swets and Zeitlinger.

Tombaugh, T. N. (1996). *Test of Memory Malingering*. Toronto, ON, Multi-Health Systems.

Table 1: Rates of passing or failing the WMT and TOMM in 144 Disability cases with heterogeneous diagnoses.

	Pass TOMM	Fail TOMM
Pass WMT effort tests	104 (72.2% of cases)	1 (0.7% of cases)
Fail WMT effort tests	25 (17.4% of cases)	14 (9.7% of cases)

Table 2: Rank order of mean scores on the WMT effort and memory subtests (percent correct) in severe traumatic brain injury and neurological patients, compared with subgroups of the Disability sample, who passed or failed WMT and TOMM.

	IR	DR	Cons.	MC	PA	DFR	LDFR
A. Pass TOMM & WMT	96.7	97.3	95.2	89.2	81.5	56.7	57.7
B. Brain injury/neurological	95.1	95.3	92.1	84.5	76.8	43.7	43.6
C. Fail WMT only	77.9	77.8	70.3	51.2	49.6	33.0	33.9
D. Fail TOMM & WMT	63.6	62.0	59.8	42.9	40.0	30.2	26.8
Rank order of groups	ABCD	ABCD	ABCD	ABCD	ABCD	ABCD	ABCD

Note: There was only one case who passed WMT but failed TOMM and that case is not shown in the table.

Table 3: Scores on averaged WMT effort and memory measures from severe traumatic brain injury and neurological patients compared with subgroups of those who passed or failed WMT and TOMM (standard deviations shown in brackets).

Average WMT Measure	Severe TBI & neurological group N = 112	Pass TOMM & pass WMT N = 104	Pass TOMM & fail WMT N = 25	Fail TOMM & fail WMT N = 14
Effort Scores	94.2%	96.4%	75.3% A,B	61.8% A,B,C
	(6.2)	(3.3)	(8.5)	(13.7)
Memory Ability	62.3%	71.3% A	41.9% A,B	35.0% A,B
	(14.9)	(13.7)	(12.5)	(8.7)

A – Significantly different from TBI and neurological group ($p < .0001$).
B – Significantly worse than patients passing TOMM and WMT ($p < .0001$).
C – Significantly worse than patients passing TOMM but failing WMT ($p < .003$).