BRAIN INJURY, 2001, VOL. 15, NO. 12, 1045–1060

Effort has a greater effect on test scores than severe brain injury in compensation claimants

PAUL GREEN†, MARTIN L. ROHLING‡, PAUL R. LEES-HALEY§ and LYLE M. ALLEN III¶

† Neurobehavioural Associates, Edmonton, Alberta, Canada
‡ Memorial Hospital at Gulfport, Gulfport, MS, USA
§ Huntsville, AL, USA
¶ CogniSyst, Inc. Durham, NC, USA

(*Received 5 March 2001: accepted 24 July 2001*)

Nine-hundred and four consecutive patients, including 80 neurological patients and 470 with head injuries, were given neuropsychological tests. All 43 test scores were converted to normative *Z*-scores and averaged, giving an Overall Test Battery Mean (OTBM). A variable measuring effort correlated 0.73 with the OTBM. The OTBM mean score was 1.20 SD lower in those who failed the Word Memory Test (WMT) than in those who passed the WMT. Sub-optimal effort suppressed the OTBM 4.5 times more than did moderate–severe brain injury. When only those making a good effort were included, patients with severe brain injuries and neurological diseases scored significantly lower than groups presumed to have no neurological impairment, but these group differences were not seen when all cases were analysed together. These data illustrate the importance of measuring and controlling for sub-optimal effort in individual neuropsychological evaluations, as well as in empirical research with similar groups of patients.

Introduction

Neuropsychologists have made statements about patient effort in clinical reports for many years. However, past estimates of effort were often based on subjective clinical impressions and effort was not routinely measured with standardized tests until recently. No corrections were made to control for error introduced by examinees' sub-optimal effort, either during individual assessments or in empirical research. It has now become apparent that subjective assessments of effort are prone to error and objective psychometric measurement of effort in forensic patients has become the standard practice [1, 2]. Many studies have shown high rates of exaggeration of cognitive impairment in certain populations, such as patients with mild head injuries claiming compensation. Binder [3] found that 33% of mild head-injured patients seeking compensation exaggerated deficits on psychometric testing. Larrabee [4] argued that the incidence of exaggeration of cognitive deficits in mild head injury patients claiming compensation was 10 times higher than the base rate for actual cognitive deficits, which implies that the majority of impaired test scores in such patients are invalid.

Correspondence to: Paul Green, PhD, Neurobehavioural Associates 201, 17107-107 Ave., Edmonton, Alberta, Canada, T5S 1G3. e-mail: paulgreen@shaw.ca

Brain Injury ISSN 0269–9052 print/ISSN 1362–301X online © 2001 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/02699050110088254

However, exaggeration is not peculiar to head injury. Cognitive exaggeration on the Amsterdam Short-Term Memory Test (ASTMT), an effort test, was evident in 61% of litigating post-whiplash patients [5]. Similar results have been found with other patient groups, including people with chronic fatigue syndrome [6], fibromyalgia [7, 8] and various other diagnoses [9]. From such findings, it would appear that the presence of financial incentives for disability is the most critical factor in determining exaggeration, rather than any particular diagnosis. Non-financial reasons for putting forth sub-optimal effort on testing have also been investigated, as illustrated by a recent report about children who freely admitted that they chose to fail effort tests and that they produced invalid results on other tests [10].

Over the years, estimates of the proportion of plaintiffs feigning psychological deficits have varied widely, from a low of 1% [11] to over 50% [12], with a possible 47% of workers' compensation patients involving malingering [13]. One study estimated the percentage of manufactured memory deficits in patients claiming persistent post-concussive syndrome as being between 33–60% [14]. An average failure rate of 30% was found on one effort test applied to 1752 compensation cases across 13 different practices in the USA and Canada [15]. Failure rates ranged from 21–76% across different sites. Therefore, when dealing with compensation seeking patients, medical disability claimants, or plaintiffs, examiners need to be aware of the high probability that some patients' test scores will be invalid due to sub-optimal effort. To aid clinical judgment, neuropsychologists would benefit from knowing how varying degrees of effort affect test scores. This requires the study of the results of many neuropsychological tests in a large number of clinical patients, in whom effort has been simultaneously measured by several methods.

There were several goals in the current study. First, one wished to measure the extent to which effort accounts for the statistical variance in neuropsychological test scores in a large clinical sample of examinees tested for purposes of determining eligibility for financial disability compensation. How big of an effect does effort have? Secondly, one was interested in the base rate of failure on effort tests when rational cut scores were applied. How many fail effort tests in each diagnostic group? Thirdly, one wished to identify the best predictor of test performance among several independent variables, including measures of effort, intelligence, age, years of education, and diagnosis. Which of these variables affects test scores the most? Finally, the analyses were designed not only to generate estimates of the percentage of variance accounted for by various measures of effort, but also to measure the degree to which sub-optimal effort suppresses examinees' test scores. How much does effort influence test scores, compared with the effects of brain injury and neurological disease? Does brain injury or neurological disease have a larger effect on test scores than effort or vice versa?

Method

Participants

Patients were seen for neuropsychological assessment as outpatients in the context of a Canadian Workers' Compensation Board claim ($n = 376$), a medical disability claim ($n = 317$) or personal injury litigation ($n = 196$). Financial benefits for disability were potentially available to or were being received by the remaining 15 patients referred privately. The sample included head injured patients ($n = 470$)

and neurological patients ($n = 80$). Neurological patients suffered from a variety of disorders, including strokes ($n = 21$), aneurysms ($n = 15$), multiple sclerosis ($n = 11$), tumour ($n = 8$), epilepsy ($n = 3$), or other miscellaneous conditions ($n = 17$; e.g. herpes simplex encephalitis, Von Hippel-Lindau disease, hypoxic event, abscess, venous thrombosis, dorsal midbrain haemorrhage).

In addition, psychiatric patients were studied ($n = 107$), including patients referred for major depression ($n = 79$), anxiety disorders ($n = 16$), bipolar mood disorders ($n = 8$), and psychotic illnesses ($n = 4$). Finally, 246 medical patients were studied, including patients with orthopaedic injuries ($n = 77$), chronic fatigue syndrome ($n = 34$), chronic pain syndrome or fibromyalgia ($n = 59$), and other various conditions ($n = 77$).

Objective measures of head injury severity

In the 470 head injury referrals, all available details of head injury severity were recorded, including the lowest Glasgow Coma Scale scores (GCS) within 24 hours of injury, the presence or absence of intracranial CT or MRI brain abnormalities, the duration of post-traumatic amnesia (PTA), and the duration of loss of consciousness (LOC). Patients with head injuries were divided into three levels based on their GCS, as shown in table 3. There were 170 patients with a GCS of 14–15 ($M = 14.8$, SD = 0.4); 22 patients with a GCS between 9–13 ($M = 11.2$, SD = 1.5) and 32 patients with a GCS between 3–8 ($M = 5.0$, SD = 1.8). If no GCS was recorded in the file, as in the case of patients who did not consult a doctor on the day of the accident, it was assumed to be 15, when there was no evidence that a patient lost consciousness, suffered any post-traumatic amnesia, nor exhibited any radiological brain abnormalities.

There were 160 patients with a head injury with no CT or MRI abnormality and 134 patients with findings of abnormality on either CT, MRI scan, or both. Other patients were not given a CT nor MRI scan. In the neurological patient group, there were reports of CT or MRI findings in 66 patients and abnormalities were present in 61 of these patients (92%).

There were 276 head injury patients with PTA less than 24 hours ($M = 0.7$, SD = 2.5, and Md = 0), and 90 patients with PTA greater than or equal to a day ($M = 360$, SD = 514.0, and Md = 168). Self-reports of PTA were not accepted, unless they were independently confirmed by previous medical reports written shortly after the accident. Such information could not be obtained about PTA in 104 patients. When the emergency room notes indicated some unspecified but short duration of amnesia, estimates of PTA were based partly on medical reports, partly on self-reports of the accident and immediate consequences on comprehensive interviewing, and partly on reports of relatives who were with the patient shortly after the accident. There were 300 patients with LOC less than 0.5 hours ($M = 0$, SD = 0.1, and Md = 0) and 44 patients with LOC greater than or equal to 0.5 hours ($M = 153.5$, SD = 256.0, and Md = 31.5). Positive LOC was rated as present only based on records from emergency medical technicians at the scene or emergency room reports and never on self-report alone, except when there was no evidence of any loss of consciousness and no other evidence of brain injury. When the patient reported accurate recall of events just before and immediately following the accident, PTA and LOC were rated as 0 hours.

There were significant Spearman's Rhos between each of the four separate criteria for head injury severity. The GCS correlated −0.88 with estimated PTA, −0.69 with LOC, and −0.57 with the presence or absence of intracranial abnormalities on CT or MRI of the brain. LOC correlated 0.76 with PTA. LOC and PTA correlated with CT/MRI abnormalities at 0.40 and 0.58.

Demographics

For the 470 patients with head injuries, the mean age was 39.0 (SD 12.1), mean years of education was 11.9 (SD 2.8) and 75% were men. For the 80 neurological patients, the mean age was 46.5 (SD 6.3), mean years of education was 13.4 (SD 3.6) and 57% were men. In all remaining diagnostic categories ($n = 354$), the mean age was 44.0 (SD 10.6), mean years of education was 12.8 (SD 3.0) and 45% were men. While there were some small but significant differences between these three major groups on these variables, they could not explain any of the major findings below. For example, the neurological patients were older than the non head-injury patients but, as shown below, they had the lowest failure rate on effort tests. Also, differences between these groups on all the effort measures were in the order of one or two percentage points, the largest being 3.4% (WMT consistency). It was the neurological patients who obtained the highest score on this measure. As seen in table 2, the effects of variables such as age, years of education and gender on the neuropsychological test scores were very greatly overshadowed by the effects of effort. In those who passed the effort tests, there were no significant correlations between years of education and CARB ($r = 0.04$), WMTIR ($r = 0.02$), WMTDR ($r = 0.06$) or WMT consistency ($r = 0.06$). Age did not correlate with WMTDR ($r = 0.06$) or CARB ($r = 0.05$) and the correlations were small between age versus WMTIR ($r = -0.13$) and WMT consistency ($r = -0.13$). Gender did not correlate significantly with any of the effort measures.

Independent measures

Nine-hundred and four patients, referred consecutively to a neuropsychologist (PG) for disability evaluations, were given two Symptom Validity Tests (SVT), designed to detect sub-optimal effort, the Computerized Assessment of Response Bias (CARB) [9] and the WMT [16–18]. These tests yielded four SVT measures, including the total score for all three blocks on the CARB, the WMT Immediate Recognition score (WMT-IR), Delayed Recognition score (WMT-DR) and Consistency score (WMT-Cons1). Each patient was also given up to 43 neuropsychological tests, as shown in table 1, one of which was the California Verbal Learning Test (CVLT) [19]. A fifth measure of effort, the CVLT Logit formula, was calculated from the results of the latter test, using the formula of Millis [20]. All patients were given 2 full days for testing and interviewing. Some patients were extremely slow and/or uncooperative, and so not all tests could be administered to all patients, and the mean number of tests given per patient was 34.

For each patient, all test results were converted to Z-scores relative to external normative data, such as those of Heaton *et al.* [21]. A single index was then generated to represent the mean Z-score for the person's average performance across all measures. This was the OTBM developed by Miller and Rohling [22]. The tests were also clustered into the six domains shown in table 1, such as execu-

Table 1. Forty-three ability measures contributing to the Overall Test Battery Mean (OTBM), grouped by domain, and five effort measures contributing to the Symptom Validity variable (SV)

43 ability measures

- *Executive Functioning*, EF ($n = 6$) Wisconsin Card Sorting Test-Categories achieved and Perseverative errors; Category Test—Errors; Thurstone Word Fluency; Ruff Figural Fluency—Total score and perseverations; Gorham's Proverbs.
- *Memory and Learning*, ML ($n = 15$) California Verbal Learning Test—Total, Trial 5, Short Delay Free Recall, Long Delay Free Recall, Recognition hits; Cognisyst Story Recall Test—Immediate and Delayed Recall; Word Memory Test—Paired Associates, Multiple Choice, Delayed recall, Long delayed recall; Rey Complex Figure Test—Delayed Recall and Recognition; Warrington Recognition Memory Test—Words & Faces.
- *Verbal Comprehension*, VC ($n = 4$) Wechsler Adult Intelligence Scale—Revised Verbal Intelligence Quotient or Multidimensional Aptitude Battery Verbal Intelligence Quotient, Wide Range Achievement Test-III—Reading, Spelling & Arithmetic.
- *Attention & Working Memory*, AW ($n = 8$) Trail Making Test—Forms A & B; Digit Span—Forward and Backward; Visual Memory Span—Forward and Backward; California Verbal Learning Test—Trial 1 and List B.
- *Perceptual Organization*, PO ($n = 4$) Rey Complex Figure Test—Copy and Recall; Benton's Judgment of Line Orientation; Wechsler Adult Intelligence Scale—Revised, Performance Intelligence Quotient.
- *Psychomotor Skills*, PS ($n = 6$) Finger Tapping—Dominant and Non-dominant; Grip Strength—Dominant and Non-dominant, Grooved Pegboard—Dominant and Non-dominant.

5 Symptom Validity Measures (SV)

- Computerized Assessment of Response Bias ($n = 1$) Total score for all three blocks of trials.
- Word Memory Test ($n = 3$): Immediate Recognition trial (IR).
- 30-Minute Delayed Recognition trial (DR).
- Consistency of responding between IR and DR.
- California Verbal Learning Test Logit formula ($n = 1$).

tive function tests or memory and learning tests. Within each domain, the person's scores were averaged, yielding six domain scores per person. A Symptom Validity composite Z-score (SV) was calculated from the five effort measures derived from the CARB, the WMT and the CVLT Logit Formula [20].

Results

Percentage of variance accounted for by symptom validity

In all 904 patients combined, the correlation between the SV measure and the global ability measure defined by the OTBM was 0.70 (Spearman's Rho) or 0.74 (Pearson's r). Hence, the composite effort measure explained between 49–54% of the variance in all test scores as reflected in the OTBM. The SV composite index accounted for more variance than any other single domain and far more than age, education, gender or any index of severity of neurological impairment in the head injured patients, as shown in table 2.

The correlations between the OTBM and each of the individual components of the SV domain score were as follows (Spearman's Rho and Pearson's r, in that order): WMT-IR $= 0.62$ and 0.66; WMT-DR $= 0.64$ and 0.69; WMT-Cons1 $= 0.66$ and 0.66; CVLT-Logit $= 0.49$ and 0.59, and CARB $= 0.44$ and 0.57. The mean of the three WMT measures correlated at 0.70 with the OTBM (Pearson's r) and 0.67 (Spearman's Rho). Because the WMT measures were the best

Table 2. Per cent variance in OTBM accounted for by each domain included in the OTBM and by demographic variables

Rank	Domain assessed	% Variance	Rank	Variable	% Variance
1	SV	53%	8	Years of education	11%
2	AW	52%	9	Age in years	4%
3	ML	49%	10	PTA (retrospective)	1%
4	PO	49%	11	LOC	1%
5	EF	41%	12	GCS	1%
6	VC	32%	13	Sex	<1%
7	PS	17%	14	Positive CT or MRI	<1%

predictors of OTBM scores and to simplify data presentation, groups demonstrating sub-optimal effort were defined as failing any of the three WMT effort measures, according to the criteria defined in the WMT manual [16].

In the whole sample, combining all diagnoses together, there was a highly significant difference on the OTBM mean score between those who passed the WMT effort measures ($M = -0.24$, SD $= 0.64$, $n = 694$) and those who failed ($M = -1.48$; SD $= 0.90$, $n = 188$; $F(1, 880) = 484$, $p < 0.0005$). There was also a highly significant difference on the OTBM mean score between those who passed the CARB ($M = -0.31$, SD $= 0.70$, $n = 701$) and those who failed ($M = -1.29$; SD $= 1.00$, $n = 185$; $F(1, 884) = 221$, $p < 0.0005$). Thus, when sub-optimal effort was defined by WMT, those with sub-optimal effort scored 1.24 SD lower on the OTBM than those who passed the effort tests. When defined by CARB, the difference in the OTBM between adequate and poor effort patients was 0.98 SD. The mean difference between those who passed or failed the CVLT-Logit was 0.98.

A combination of CARB and WMT to define sub-groups revealed that 685 patients passed both the WMT and the CARB and their mean OTBM was −0.26; 31 patients failed CARB only (OTBM $M = -0.65$); 82 patients failed WMT only (OTBM $M = -1.20$) and 103 patients failed both CARB and WMT (OTBM $M = -1.66$). The differences between the latter OTBM scores were significant ($F(3, 897) = 129.0$, $p < 0.0005$). The mean OTBM was significantly higher in those who passed both tests than in those who failed the WMT or who failed both CARB and WMT (Bonferroni $p < 0.002$), but there was no significant difference in the OTBM between those who failed only the CARB and those who passed both CARB and WMT.

Worse than chance performances, defined as a score of less than 50% correct, were observed in the following percentages of cases for each test: WMTDR (4.4%) CARB (1.8%), Warrington Recognition Memory Test (RMT) Words (1.7%), Warrington RMT Faces (2.7%), CVLT Recognition Hits (3.4%). Those who failed the WMT were divided further into those who scored at a worse than chance level on any of the latter tests ($n = 78$) and those who did not score at a worse than chance level ($n = 145$). The mean OTBM in the worse than chance scorers (−2.1, SD 1.0) was significantly lower than the mean OTBM score from those who failed WMT but who did not produce any worse than chance scores on the above tests (−1.0, SD 0.7), $F = 73$, df (1,221), $p < 0.0005$. In those who passed the WMT, the mean OTBM was −0.22 (SD 0.63). Hence, compared with those passing the WMT, the OTBM was suppressed by a mean of 0.78 SD in WMT failures with no worse than chance scores and by 1.88 SD in those with at least one worse than chance score. Because a worse than chance score on any test is widely regarded as the strongest evidence of deliberate exaggeration, it is interesting to note that the following mean scores were observed in those scoring at a worse than chance level on at least one test: CARB total score ($M = 73\%$, SD 24), WMT IR ($M = 61\%$, SD 20), WMT DR ($M = 57.4\%$, SD 19.8), WMT consistency IR to DR ($M = 63\%$, SD 15.8), CVLT recognition hits ($M = 8.7$, SD 3.7) Warrington RMT Words ($M = 28.3$, SD 8.3) and Warrington RMT Faces ($M = 27.2$, SD 8.3). All of the latter mean scores are above 50% correct, but they were associated with a mean suppression of 1.88 SD in scores on the OTBM, derived from 43 neuropsychological tests.

Variance accounted for by referral source

Of the 904 patients in this study, 28.5% of patients failed one or more of the WMT effort measures and 22.0% failed the main effort measure, WMT-DR. There was a significant difference between rates of WMT failure dependent upon referral source ($\chi^2 = 18.4$, $df = 6$, $p < 0.005$), using failing any of the three WMT effort measures as the criterion. The highest failure rate was 35% in Workers' Compensation Board referrals, who were predominantly patients with mild head injuries. The failure rate was 24.6% for people involved in personal injury litigation and 23.0% for medical disability insurance claims through insurance companies or Government agencies (all other diagnoses).

Variance accounted for by diagnosis

There was a significant difference between the rates of failure by diagnosis. The more severe the impairment objectively, the lower the failure rates on WMT. The failure rates on the WMT effort measures were 16.0% in 80 neurological patients, 33.0% in 470 head injury patients, 28.0% in 78 patients with a diagnosis of major depression and 25.0% in 276 patients with all other diagnoses combined, Kruskall-Wallace, $\chi^2(3) = 11.7$, $p = 0.008$. Of the 276 head injury patients with PTA less than 24 hours, 34% failed the WMT, whereas the failure rate was only 18% for the 90 patients with PTA of 1 day or more, Mann Whitney $Z = -2.86$, $p = 0.004$. The 16% WMT failure rate in the patients with neurological diseases was significantly lower than the 34% failure rate in head injury patients with less than 24 hours PTA, Mann Whitney $Z = -3.00$, $p = 0.003$. The neurological patients making an adequate effort had a mean OTBM of -0.40 (SD $= 0.60$), which was not significantly different from the mean OTBM of -0.35 (SD $= 0.68$) in the head injury patients with PTA of more than 1 day, who passed the WMT effort measures, $F(1, 181) = 0.25$, $p = 0.60$.

By diagnostic class, the mean OTBM scores for those passing and failing the WMT were, respectively, as follows: head injuries with PTA less than 24 hours (-0.18, vs. -1.34); head injuries with PTA 1 day or more (-0.35 vs. -1.47); neurological patients (-0.40 vs. -1.62); orthopaedic injuries (-0.34 vs. -1.70); chronic fatigue syndrome (-0.04 vs. -1.64); chronic pain syndrome (-0.31 vs. -1.65); major depression (-0.03 vs. -1.45) and other diagnoses (-0.30 vs. -1.53).

Magnitude of suppression due to head injury severity

In tables 3 and 4, patients with head injuries are broken down into different levels of head injury severity, according to four separate conventional criteria: (1) GCS, (2) abnormalities on a CT or MRI, (3) PTA duration, and (4) duration of LOC. The bottom line of table 4 shows that those who failed the WMT (exaggerators) scored between 1.18–1.24 SD lower on the OTBM, compared with patients who passed the WMT test (genuine). In contrast, the mean difference in the OTBM between the genuine mild head injuries versus the genuine severe brain injury cases was only 0.27 SD. Hence, the mean degree of OTBM suppression by sub-optimal effort (1.21 SD) was, on average, 4.5 times greater than the difference between the OTBM scores in people with mild versus severe brain injuries (0.27 SD). In those who failed the WMT, suppression of the OTBM occurred to approximately the same degree, whether the head injury was mild or more severe.

In patients who passed the WMT effort subtests, those with a relatively mild head injury scored, on average, 0.12 SD below the normal mean (see table 4), whereas exaggerators with mild head injuries scored 1.36 SD below the normal mean. Thus, sub-optimal effort led to the equivalent of a mean drop in the OTBM score of 1.24 SD in the mild head injury patients. In contrast, in the genuine patients with more severe head injuries, based on the four different criteria in tables 3 and 4, the OTBM was only 0.39 SD below normal. In those passing the WMT, the most severely head injured patients scored, on average, 0.27 SD lower than the genuine patients with mild head injuries.

WMT failure was inversely related to degree of radiological brain abnormality. In head injured patients who passed the WMT test, there were CT or MRI abnormalities of the brain in 53% of cases, which was significantly greater than the 28.5% incidence of CT or MRI abnormalities in patients who failed the WMT, $F(1{,}285) = 14.6$, $p < 0.0005$. In the remaining patients, no CT or MRI had been done or, if so, the reports were unavailable. There were no differences in terms of mean GCS scores between the patients who passed the WMT ($M = 13.0$, $SD = 3.6$) and patients who failed the WMT ($M = 13.1$, $SD = 3.8$), $F(1{,}215) = 0.06$, $p = 0.8$. Similarly, there was no difference in the mean PTA

Table 3. Overall Test Battery Mean (OTBM) scores for each level of effort and for each level of head injury severity, defined by Glasgow Coma Scale score (GCS) or Post-Traumatic Amnesia duration (PTA)

Levels of effort based on pass/fail WMT	OTBM scores by levels of head injury severity			
	GCS 3–8, ($n = 32$)	GCS 9–13, ($n = 22$)	GCS 14–15, ($n = 170$)	Difference (most vs. least severe injuries)
PASS (P)	−0.47	−0.43	−0.13	0.34
FAIL (F)	−1.78	−1.40	−1.37	0.41
P–F difference	−1.31	−0.97	−1.24	
	PTA $\geq$ 1 day ($n = 90$)		PTA<1 day ($n = 276$)	
PASS (P)	−0.30		−0.13	0.17
FAIL (F)	−1.57		−1.35	0.22
P–F difference	−1.27		−1.22	

Table 4. Overall Test Battery Mean (OTBM) scores for each level of effort and for each level of head injury severity, defined by the presence or absence of CT/MRI abnormalities or by duration of LOC. Overall results shown at end of table

	OTBM scores by level of head injury severity		
Levels of effort based on pass/fail WMT	CT/MRI abnormal ($n = 134$)	CT/MRI normal ($n = 160$)	Difference (most vs. least severe injuries)
PASS (P)	−0.33	−0.07	0.26
FAIL (F)	−1.39	−1.36	0.03
P–F difference	−1.06	−1.29	
	LOC > 0.5 hours, ($n = 44$)	LOC < 0.5 hours, ($n = 300$)	
PASS (P)	−0.48	−0.14	0.34
FAIL (F)	−1.53	−1.36	0.17
P–F difference	−1.10	−1.22	
Average values	Most severe head injuries	Least severe head injuries	
PASS (P)	−0.39 SD	−0.12 SD	0.27 SD
FAIL (F)	−1.57 SD	−1.36 SD	0.21 SD
P–F difference	−1.18 SD	−1.24 SD	

between the patients who passed the WMT ($M = 98$ hours, SD $= 310$) and patients who failed the WMT ($M = 68$ hours, SD $= 261$), $F(1,357) = 0.64$, $p = 0.42$. Finally, there was no difference in the mean duration of LOC between patients who passed the WMT ($M = 19.5$, SD $= 95.0$) and patients who failed the WMT ($M = 22.3$, SD $= 146.0$), $F(1,336) = 0.05$, $p = 0.80$.

In the neurological patient group, there were 67 patients who passed the WMT subtests and 13 patients who failed. Their respective OTBM scores were −0.40 and −1.62, $F(1,74) = 42.8$, $p < 0.0005$. The difference is 1.22 SD, a magnitude equivalent to the differences observed between good and poor effort cases in the more severely brain injured patients (1.18), the mild head injured patients (1.24), and in the non-head-injury and non-neurological group (1.35). Sub-optimal effort, therefore, has roughly equivalent effects on test scores in all the diagnostic groups examined. Exaggeration produced an equivalent drop in mean scores of between 1.18–1.35 SD, compared with the adequate effort patients.

Comparisons between three main groups

To analyse the effects of effort on the OTBM further, three main groups were created, which were:

(1) all patients with known or probable cerebral impairment, consisting of the neurological patients ($n = 80$) and the moderate-to-severe brain injury patients, who were selected for PTA of 1 day or more or a GCS less than 13 ($n = 151$). There were abnormal CT scans in 88% of the latter patients. The mean PTA in the head injured patients was 254 hours (SD $= 459$,

median = 72 hours), their mean LOC was 67 hours (SD = 184, median = 0.30 hours), and their mean GCS was 10.4 (SD = 4.4, median = 12). This was called the TBI-NEURO group ($n = 252$);

(2) the least severe head injury patients, with PTA of less than 1 day and who did not have abnormal CT scans ($n = 271$). Their median PTA was zero (M PTA = 0.4 hours, SD = 1.5), their median LOC was zero ($M = 0.03$ hours, SD = 0.14) and their median GCS was 15 (M GCS = 14.8, SD = 0.4). This was called the mild head injury group (MHI); and

(3) all patients in the other diagnostic groups ($n = 341$), which was called the miscellaneous or MISCEL group.

The TBI-NEURO group was expected to score lower on the OTBM than the patients in the other two main groups. However, table 5 (step 1) shows that, when all subjects from these three groups were included, there was no significant difference on the OTBM between the TBI-NEURO group, the MHI group, and the MISCEL group, $F(2{,}898) = 0.60$, $p = 0.54$. However, there was a significant difference amongst these groups on the two effort measures applied, which were the SV composite score, $Z(2) = 6.5$, $p = 0.04$, and the delayed recognition measure of the WMT, $Z(2) = 7.0$, $p = 0.03$. For the latter two comparisons, the Mann Whitney test was used, because of the non-normal distributions of these variables. Then, all patients suspected of producing invalid results because they failed the WMT (table 5, step 2) were removed. Then, there was a significant difference between the groups on the OTBM in the expected direction. The TBI-NEURO group ($n = 186$) scored significantly lower than the MHI ($n = 183$) and the MISCEL group ($n = 315$; $F(2{,}691) = 4.7$, $p = 0.009$).

In the third step shown in table 5, all patients from the three groups who failed the WMT-DR were compared, and these groups did not differ from each other on

Table 5. Differences between OTBM scores and symptom validity scores in three main groups from 904 patients

	OTBM	SV	WMT-DR
Step 1: All patients			
TBI-NEURO	−0.57 (0.8)	−0.74 (1.8)	−1.4 (3.2)
MHI	−0.52 (0.8)	−1.4 (2.5)	−2.5 (4.2)
MISCEL	−0.9 (0.9)	−1.1 (2.4)*	−1.9 (4.1)*
Step 2: Genuine patients only			
TBI-NEURO	−0.37 (0.6)	−0.04 (0.7)	−0.09 (1.1)
MHI	−0.18 (0.6)	−0.12 (0.9)	−0.23 (1.3)
MISCEL	−0.21 (0.7)**	−0.03 (0.8)	−0.05 (1.1)
Step 3: Exaggerators only			
TBI-NEURO	−1.5 (0.7)	−4.1 (2.0)	−7.5 (3.2)
MHI	−1.3 (0.9)	−4.7 (2.1)	−8.3 (3.5)
MISCEL	−1.6 (0.9)	−5.3 (2.2)*	−9.1 (3.7)*
Step 4: Genuine TBI-NEURO patients compared with exaggerators in other groups			
TBI-NEURO (gen)	−0.37 (0.6)	−0.04 (0.7)	−0.09 (1.1)
MHI (exagg)	−1.34 (0.9)	−4.7 (2.1)	−8.3 (3.5)
MISCEL (exagg)	−1.59 (0.9)***	−5.3 (2.2)***	−9.1 (3.7)***

Differences between three groups: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the OTBM, $F(2, 185) = 1.76, p = 0.17$ (TBI-NEURO $n = 38$, MHI $n = 73$, and MISCEL $n = 77$). There were significant differences on the SV measure (Kruskall Wallace, $\chi^2(2) = 7.7$, $p < 0.03$) and on the WMT effort measure ($\chi^2(2) = 6.3$, $p < 0.04$), with the TBI-NEURO scoring higher than the other two groups. In the fourth step in table 5, the genuine TBI-NEURO patients, defined by passing WMT-DR, were compared with the exaggerators from the other two groups. The difference was highly significant and in favour of higher performances by the TBI-NEURO patients, $F(2,333) = 91.3$, $p < 0.0005$. This was contrary to expectation, based on diagnosis, but not contrary to expectation based on WMT scores.

Effort as a continuous quantitative variable

In the current sample, a gradient of effort was created by classifying levels of scores on the mean of the WMT-IR, WMT-DR and WMT-Cons1 effort measures, relative to the mean scores from patients with moderate-to-severe brain injuries and with a mean GCS of 9, as described in detail by Allen and Green [24]. With the brain injured mean as zero, the ranges of scores on the WMT and the sample sizes for the groups shown in figure 1 are as follows,: scores above −1 SD ($n = 562$), scores −1 to −2.9 SD ($n = 126$), −3 to −4.9 SD ($n = 83$), −5 to −6.9 SD ($n = 54$) and scores at or lower than −7 SD ($n = 76$). The graph shows a steady decrease in the OTBM as the WMT effort scores decrease.

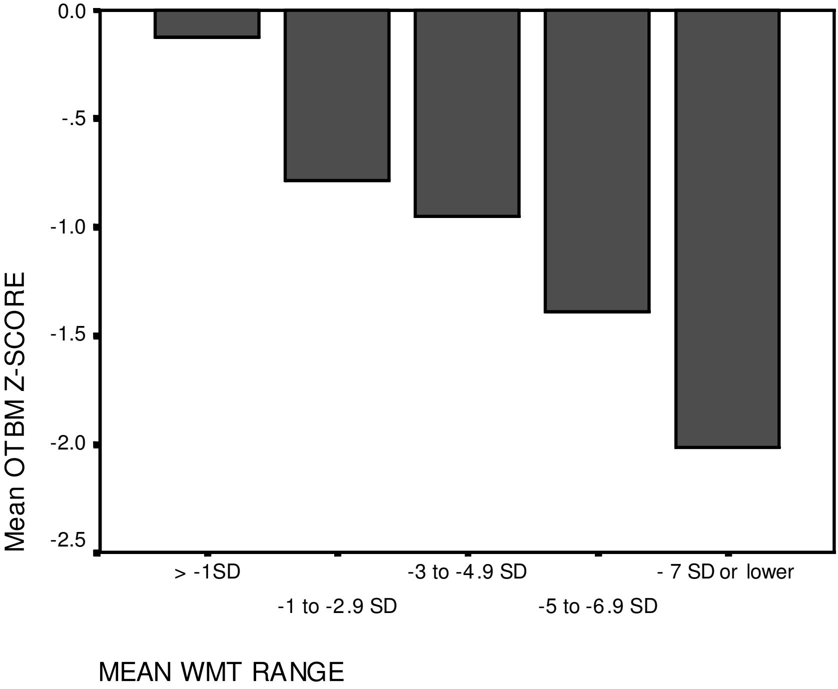

Figure 1. Ranges of scores on the mean of three WMT effort measures (IR, DR and Consistency), expressed as Z-scores relative to the mean from brain injured patients, and the corresponding gradient of scores on the OTBM.

Neurocognitive domains and specific instruments

Considering all 904 patients, the differences in each of the domain scores between those passing and failing the WMT effort measures were highly significant in all comparisons ($p < 0.0005$). The mean scores in each domain for those passing and failing the WMT effort measures were, respectively, as follows (ranked from the most affected by effort to the least affected): Memory and Learning (−0.31 vs. −2.17, difference = 1.86, $n = 694$ and 188); Attention and Working Memory (−0.23 vs. −1.46, difference = 1.23, $n = 646$ and 170); Perceptual Organization (−0.14 vs. −1.20, difference = 1.06, $n = 606$ and 161); Psychomotor Skills (−0.35 vs. −1.29, difference = 0.94, $n = 500$ and 132); Verbal Comprehension (−0.30 vs. −0.87, difference = 0.57, $n = 669$ and 174); Executive Functions (−0.45 vs. −1.19, difference = 0.26, $n = 601$ and 156). Thus, sub-optimal effort affected tests of memory and learning (ML) the most and executive functions (EF) the least.

In the patients with head injuries and PTA of 1 day or more, who passed the WMT effort tests, the mean Memory and Learning (ML) score was −0.46 (SD = 0.90) and in those who failed the WMT, it was −2.26 (SD = 1.00). In the patients with head injuries and no PTA or PTA of less than 1 day, who passed effort tests, the mean ML score was −0.26 (SD = 0.90) and in those who failed the WMT, it was −1.97 (SD = 1.30). Thus, the mean ML score in the least severe head injury patients failing WMT was 1.50 SD lower than the mean score from those with the most severe head injuries who passed the WMT.

Discussion

It has previously been reported that groups of patients with mild head injuries scored significantly lower on the WMT effort tests and on CARB than groups of people with more severe brain injuries [17, 23]. There was a greater proportion of exaggerators in the mild head injury groups. In the present study, the MHI group also scored significantly lower on the symptom validity measures than the more severely injured patients in the TBI-NEURO group (table 5, Step 1, SV and WMT measures). Similarly, in this study, patients with moderate-to-severe brain injuries or neurological diseases failed effort tests *less* often than the group composed of people with mild head injuries, major depression, orthopaedic injuries, chronic fatigue syndrome and chronic pain.

When people fail effort tests, it is usually assumed that their other test results are likely to be invalid. This study went one step further and proved that they were invalid. The authors were not only able to measure the presence of exaggeration with effort tests but also to study the effects of such exaggeration on a wide range of neuropsychological test scores, represented by the OTBM. This global measure was based on a total of 30,736 individual test scores (904 patients with a mean of 34 test scores per patient). Effort explained 53% of the variance in these data. To put this finding in perspective, years of education is a variable thought to have a significant impact on ability test scores, but it explained only 11% of the variance in the same OTBM data, and age explained only 4% of the variance.

The OTBM allowed one to compare patients' scores with the means from normative samples of community living adults, whose performance level would be represented by a score of 0 on the OTBM. Figure 1 shows how strongly scores on an effort test, the WMT, predicted scores on the OTBM. The graph shows that effort varies along a continuum and that, as effort reduces (or as exaggeration increases), the patients' scores on many neuropsychological tests decline sharply and severely. In the graph, in all but the highest scoring group on the WMT, the mean OTBM was lower than the mean score from patients with severe brain injuries who passed the WMT. The domain most affected by effort was 'memory and learning' and, in this domain, the mean score of those failing WMT effort measures was 1.86 SD below the mean from those passing them.

The effect of effort was sufficient to turn expected group differences upside down. For example, one would expect to find that both the MHI and the MISCEL group would score higher on the OTBM than the more severely brain injured and neurological patients. In fact, they did not. When all patients were studied together, the expected effect of more severe brain injury and neurological disease was not evident in a lower OTBM in the TBI-NEURO group than in the other two groups. However, these groups included invalid test results from some cases who failed effort tests. When only those patients who had passed the WMT effort measures were studied, there was a highly significant impairment of test scores on the OTBM in the TBI-NEURO group, compared with the relatively mild head injury patients (MHI) and patients of other diagnoses (MISCEL; table 5 step 2).

Selecting different subjects, one was able to show just the opposite between-group differences. In table 5 (step 4, OTBM), the members of the TBI-NEURO group who passed the WMT effort measures scored significantly higher on the OTBM than patients from the other two groups who failed the effort tests. The mean OTBM score in the mild head injury group who failed the effort tests was −1.34; meaning that, on up to 43 tests, this group's mean score was, on average, 1.34 SD below normal. In IQ scores, this would be equivalent to mild head injury patients having an IQ of 81. In contrast, the true effect of moderate-to-severe brain injury and neurological disease can be seen in a mean OTBM of only −0.37, representing a mean score from 43 tests which was 0.37 SD below normal (table 5, step 4). This is equivalent to an IQ score of 94.5. Using the OTBM scores as the metric, people with mild head injury who were assessed as part of a compensation claim and who failed effort tests, scored 3.6 times further below the normal mean score than those in the known cerebral impairment group (STBI-NEURO). Hence, exaggeration has a far bigger effect on test scores than brain injury or disease. In the same way, the bottom line on table 4 shows that the patients with the least severe head injuries, who failed the effort tests, scored a mean of 1.36 SD below normal on the OTBM, which was, on average, more than 10 times further below the normal mean than the mean score from the mild head injury patients who passed the effort tests (mean OTBM = −0.12). Very similar results were found whether head injury severity was defined by LOC, GCS, PTA, or CT abnormalities, and so these findings are very robust.

Some patients from all diagnostic groups failed effort tests. What is notable is that the failure rates on effort tests were the lowest in the groups known to have the most severe cognitive impairment (i.e. the most severe brain injury patients and the neurological patients). When patients were exaggerating their cognitive difficulties, their OTBM scores were extremely low, regardless of diagnosis. The mean OTBM scores from exaggerators in different diagnostic groups was −1.57, whereas patients, who were unresponsive to verbal commands for 14–28 days after head injury [25] had a mean OTBM score of −1.33 [26]. On the other hand, in the genuine patients in the present study, combining all diagnoses, the mean OTBM score was only 0.20 SD below normal.

If such sources of error are not removed from clinical test data, it will result in illogical conclusions that are incompatible with neurological reality. For example, patients with mild head injuries or major depression, who fail effort tests, might be reported as suffering from more cognitive impairment and, therefore, as deserving greater financial compensation than patients with severe brain injuries, who make a full effort and produce valid test results. This would represent a major injustice to the more severely injured patients and it would mean that exaggerated deficits were being selectively rewarded. The implications of these findings are, therefore, clear. It is essential to measure effort routinely in any clinical neuropsychological assessment, at least when the test results are to be used to evaluate eligibility for financial compensation. In fact, now that it has been found that the effects of effort on test results can be several times greater than the effects of severe brain injury in all the diagnostic groups studied, it would probably be advisable to control for sub-optimal effort in all examinations, until it becomes clear from empirical data that variations in effort are not contaminating data from a particular population. If we simply assume that test results are valid and do not measure effort, we run the risk of failing to control for a variable that appears to have more of an influence on test scores than age, education, intelligence, abnormal CT findings, extensive PTA, extended LOC and a low GCS all added together, as shown in the variance figures in table 2.

Significant effects of sub-optimal effort in suppressing test scores have been reported in both adults and children, for whom there were no financial incentives [10]. There are many reasons, apart from money, why some people might put forth inadequate effort during testing. This has been recognized for generations by the fact that, in any clinical assessment, statements have traditionally been made about the amount of effort put forth by patients, regardless of whether there were any financial incentives for exaggeration. Such statements implicitly acknowledge the fact that less than complete effort is a real possibility in all cases and that test results can be rendered invalid by sub-optimal effort.

The current findings show that undetected invalid effort could lead to false conclusions in clinical research with groups of patients. Therefore, it might be advisable to include measures of effort in virtually all research studies, especially when the patients' illnesses or injuries are potentially sufficient to lead to disability and, therefore, in most patients, to possible compensation for disability. When exaggeration in mild head injury patients can lead to much greater apparent impairment in test scores than severe brain injury, as shown above, it is not reasonable to assume that test results from mild head injury patients are valid without formally ruling out sub-optimal effort and the associated invalid test results. In fact, sub-optimal effort on the WMT was even found in 18% of patients with the most severe brain injuries and in 16% of patients with neurological diseases in the current study. Whereas the mean OTBM for all neurological patients was −0.64, it was only −0.40 when those failing the WMT were excluded. The corresponding figures for 32 severe brain injury patients with a GCS less than 9 were −0.71 (for all patients) and −0.43 (for those passing the WMT). A small percentage of patients who are making a sub-optimal effort can lead to an inflated estimate of the effects on test scores of a neurological disease or a severe brain injury. Yet, effort has typically not been controlled and has been assumed to be adequate, as in the studies of Dikmen *et al.* [25] and Volbrecht *et al.* [27].

The OTBM from groups of patients drawn from both of the latter studies have been reported [26]. The most minor head injury group in the Volbrecht *et al.* [27] study had an OTBM of −0.32 and the most severely brain injured group's mean OTBM was −1.74. Although Volbrecht *et al.* stated that a forced choice effort test was used with all patients in the study, the paper reported no results from the effort test and did not state how neuropsychological test scores were affected by effort, as measured by this instrument. No patients were excluded from the study because of low effort test scores. In view of the high base rate for exaggeration in the current study of people eligible for disability payments and in many other studies [4], it may be assumed that test scores would have been suppressed by sub-optimal effort in at least some of the patients of Volbrecht *et al.* [27], bringing down the group means on test scores. This suggestion is supported by the fact that, in the latter study, the most mild head injury group obtained a mean OTBM of −0.32, whereas the most mild head injury patients in the current study scored only −0.12 on a very similarly constructed OTBM. The OTBM in the mildest head injury group of Dikmen *et al.* [25] was −0.02 or almost identical to the normal mean and their scores were not different from those of non head-injured orthopaedic controls. In addition, the more severe brain injury patients with GCS scores of 8 or less, although perhaps not as severe as some of the patients of Volbrecht *et al.* [27], scored only −0.43 on the OTBM, which is not very different from the −0.32 mean from Volbrecht's most mild head injury patients.

These comparisons raise the possibility that sub-optimal effort has significantly influenced the results of many past studies that have not controlled for effort. They suggest that it would be desirable for effort to be measured in future studies of the effects of brain injuries on test scores and that data should be reported on the degree to which effort affects test scores. The assumption that all test results are valid in any sample is becoming increasingly implausible. There are numerous studies reporting high rates of exaggeration in patients with compensation incentives [4], but there are no studies employing adequate effort testing, which report no exaggeration at all in people eligible for compensation. The current finding of overall failure rates on the WMT of between 23–35%, depending on referral source, is consistent with estimates of the base rates of exaggeration from several studies quoted by Larrabee [4], which show that the rate of exaggeration in patients involved in compensation, disability or personal injury claims is closer to 50% than to zero. The current findings confirm the numerous studies published in the last decade attesting to the power of compensation-seeking as a confounding variable that leads to unexplained and illogically poor scores on most tests of ability. They show that effort has such a large effect that, if not controlled, it literally inverts the expected group differences between severe versus mild traumatic brain injury patients. This powerfully underscores the need for symptom validity testing, not only in forensic neuropsychological assessment, but also in group studies.

References

1. Iverson, G. L. and Binder, L. M.: Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation*, **15**: 829–858, 2000.

2. Sweet, J. J.: Malingering: differential diagnosis. In: J. J. Sweet (editor) *Forensic Neuropsychology Fundamentals and Practice* (Lisse: Swets and Zeitlinger), 1999.
3. Binder, L. M.: Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology*, **15**: 170–182, 1993.
4. Larrabee, G. J.: Neuropsychology in personal injury litigation. *Journal of Clinical and Experimental Neuropsychology*, **22**: 702–707, 2000.
5. Schmand, B., Lindeboom, J., Schagen, S. *et al.*: Cognitive complaints in patients after whiplash injury: the impact of malingering. *Journal of Neurology Neurosurgery and Psychiatry*, **64**: 339–343, 1998.
6. Van der Werf, S. P, Prins, J. B., Jongen, P. J. *et al.*: Abnormal neuropsychological findings are not necessarily a sign of cerebral impairment: a matched comparison between chronic fatigue syndrome and multiple sclerosis. *Neuropsychiatry, Neuropsychology and Behavioural Neurology*, **13**: 199–203, 2000.
7. Gervais, R., Green, P., Russell, A. S. *et al.*: Failure on symptom validity tests associated with disability incentives in fibromyalgia patients. *Archives of Clinical Neuropsychology*, **15**: 841–842, 2000.
8. Gervais, R., Russell, A. S., Green, P. *et al.*: Effort testing in fibromyalgia patients with disability incentives. *Journal of Rheumatology*, **28**: 1892–1899, 2001.
9. Allen, L., Conder, R. L., Green, P. *et al.*: *CARB' 97 Manual for the Computerized Assessment of Response Bias* (Durham, NC: CogniSyst), 1997.
10. Flaro, L. and Green, P.: Symptom validity test results with children: CARB and WMT. *Archives of Clinical Neuropsychology*, **15**: 840, 2000.
11. Keiser, L.: *The Traumatic Neurosis* (Philadelphia: J.B. Lippincott), 1968.
12. Miller, H. and Cartlidge, N.: Simulation and malingering after injuries to the brain and spinal cord. *Lancet*, **1**: 257–261, 1972.
13. Youngjohn, J. R.: Malingering of neuropsychological impairment: an assessment strategy. *A Journal for the Expert Witness, the Trial Attorney, the Trial Judge*, **4**: 29–32, 1991.
14. Greiffenstein, M. F., Baker, W. J. and Gola, T.: Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, **6**: 218–224, 1994.
15. Allen, L. M., Richards, P. M., Green, P. *et al.*: Performance patterns on the Computerized Assessment of Response Bias in 1752 compensation cases (Abstract). *Archives of Clinical Neuropsychology*, **13**: 15–16, 1997.
16. Green, P., Allen, L. and Astner, K.: *The Word Memory Test: A User's Guide to the Oral and Computer-Administered Forms, US Version 1.1* (Durham, NC: CogniSyst), 1996.
17. Green, P., Iverson, G. L. and Allen, L. M.: Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury*, **13**: 813–819, 1999.
18. Iverson, G., Green, P. and Gervais, R.: Using the Word Memory Test to detect biased responding in head injury litigation. *The Journal of Cognitive Rehabilitation*, **17**: 4–8, 1999.
19. Delis, D. C., Kramer, J. H., Kaplan, E. *et al.*: *California Verbal Learning Test: Adult Version Manual* (San Antonio, TX: The Psychological Corporation), 1987.
20. Millis, S.: CVLT Logit function (unpublished). Presented at a workshop by S. Millis and W. Van Gorp on detection of malingering at National Academy of Neuropsychology, San Antonio, TX, 1999.
21. Heaton, R. K., Grant, I. and Matthews, C. G.: *Comprehensive Norms for an Extended Halstead-Reitan battery* (Odessa, FL: Psychological Assessment Resources, Inc), 1991.
22. Miller, L. S. and Rohling, M. L.: A statistical interpretive method for neuropsychological test data. *Neuropsychology Review*, 2001, in press.
23. Green, P. and Iverson, G. L.: Validation of the computerized assessment of response bias in litigating patients with head injuries. *The Clinical Neuropsychologist*, 2001, in press.
24. Allen, L. and Green, P.: Severe TBI sample performance on CARB and the WMT. *Supplement to the CARB '97 and Word Memory Test Manuals* (Durham, NC: CogniSyst), 1999.
25. Dikmen, S. S., Machamer, J. E., Winn, H. R. *et al.*: Neuropsychological outcome at 1-year post head injury. *Neuropsychology*, **9**: 80–89, 1995.
26. Rohling, M. L., Millis, S. and Meyers, J.: Generating a linear function of residual impairment from TBI: a comparison between the HRB and a flexible battery. *Archives of Clinical Neuropsychology*, **15**: 821–822, 2000.
27. Volbrecht, M. E., Meyers, J. E. and Kaster-Bungard, J.: Neuropsychological outcome of head injury using a short battery. *Archives of Clinical Neuropsychology*, **15**: 251–265, 2000.