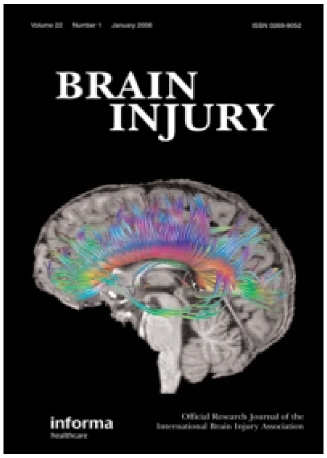

Brain Injury
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713394000

Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives

Lloyd Flaro [a]; Paul Green [a]; Ellen Robertson [a]
[a] Private Practice, Edmonton, Canada

Online Publication Date: 01 January 2007
To cite this Article: Flaro, Lloyd, Green, Paul and Robertson, Ellen (2007) 'Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives', Brain Injury, 21:4, 373 - 383
To link to this article: DOI: 10.1080/02699050701311133
URL: http://dx.doi.org/10.1080/02699050701311133

Brain Injury, April 2007; 21(4): 373–383

informa healthcare

Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives

LLOYD FLARO, PAUL GREEN*, & ELLEN ROBERTSON

Private Practice, Edmonton, Canada

(Received 12 December 2006; accepted 28 February 2007)

Abstract

Primary objective: Motivation has an important influence on neuropsychological test performances. This study examined effort on the Word Memory Test (WMT) in groups with differing external incentives.

Research design: 774 adults with Traumatic Brain Injury (TBI), tested as part of a Workers' Compensation, disability or personal injury claim stood to gain financially by appearing impaired on testing. In contrast, parents ordered by the Court to undergo a parenting assessment were highly motivated to do their best on cognitive tests because their goal was to regain custody of their children.

Outcomes and results: Consistent with these assumptions, 98.3% of 118 parents seeking child custody passed the WMT effort subtests but in cases of mild TBI the pass rate on the WMT was only 60%. The WMT failure rate in the mild TBI sample was 23 times higher than in the group of parents seeking custody. WMT failure was twice as frequent in the mild TBI group than in those with more severe TBI. WMT failure was also much higher in adults with mild TBI than in children with significant impairment from various clinical conditions.

Conclusions: Such differences in failure rates on the WMT effort subtests cannot be explained by differences in cognitive skills but they are explainable by differences in external incentives. The findings support the recommendation that objective tests of effort should be used when evaluating cognitive impairment.

Keywords: *Brain injury, motivation, effort, symptom validity test, incentive*

Introduction

Performances on neuropsychological tests are often used to make inferences regarding examinees' cognitive abilities and brain function [1]. However, neuropsychological tests can yield impaired scores indistinguishable from the effects of brain injury, which do not reflect brain-behaviour relationships and which are better explained by poor effort or exaggeration of cognitive deficits. That is why it is now recommended that objective tests of effort should be used in all neuropsychological assessments, including clinical evaluations [2, 3]. Many things can potentially affect a person's effort to do well and their engagement in the assessment. In the current study, we examined how external incentives impacted performance on effort testing by comparing groups with different external incentives, including (a) people with an external incentive to perform poorly on testing, (b) people with a strong external incentive to do well on testing and (c) people with no obvious external incentive, either to exaggerate impairment or to do well.

There have been many studies of effort testing in people with strong incentives to appear impaired and it is not unusual to find a substantial percentage of such samples failing effort tests. For example, more than 50% of prisoners facing court appearances for sentencing were judged to be malingering cognitive impairment, based partly on their failure on effort tests including the WMT [5]. Likewise, in 904 consecutive compensation cases of mixed diagnoses,

Correspondence: Dr Lloyd Flaro, Ed.D Suite 210, 17010 103 Avenue, Edmonton, Alberta, Canada, T5S 1K7. E-mail: drflaro@telus.net
*Statement of interest: Dr Paul Green is the inventor, author and copyright owner of the Word Memory Test.

ISSN 0269–9052 print/ISSN 1362–301X online © 2007 Informa UK Ltd.
DOI: 10.1080/02699050701311133

28% failed the effort measures of the WMT [6]. Additionally, Sullivan [7] reports a 48% rate of failure rate of the WMT in a series of adult university students seeking assessment for attention deficit hyperactivity disorder (ADHD), whereas only 12% of children with ADHD and other disorders failed the WMT [8]. Larrabee [9] estimated that 40% of mild head injury cases in litigation fail effort tests and, in such cases, the neuropsychological test scores would be considered invalid.

In the current study, as a group with an external incentive to appear impaired, we used a series of 774 consecutive cases that were referred for neuropsychological evaluation based on a claim of TBI. In 577 cases, the injury severity was relatively mild but it was moderate to severe in the remainder. They received the WMT as part of a comprehensive neuropsychological assessment. All these cases had an external incentive to exaggerate impairment because they were receiving disability payments or were claiming such payments from the Workers Compensation Board (WCB) or medical disability insurance companies or were seeking compensation for injuries through personal injury civil litigation.

This is the first study investigating effort in low functioning individuals with a very strong external incentive to perform well. The members in our second group were actively involved in child custody battles. We investigated performance on the effort and memory subtests of the WMT [1, 4] in 118 consecutively referred adults, who were ordered by the Alberta Family Court and the Department of Social Services to undergo an assessment of their fitness to be parents. They were referred to the first author for an assessment of parenting skills. Undergoing a fitness evaluation can be a time-consuming, arduous and potentially embarrassing situation and seeking child custody is optional and not a mandatory undertaking. Good performance on cognitive tests would maximize their opportunity to regain custody of their children. Therefore, it may be assumed that the parents seeking child custody were, in most cases, motivated to do well and we would expect a low failure rate on effort testing in this group arising from poor effort.

The third group consisted of people with little or no incentive either to exaggerate impairment or to do well. It is difficult to identify groups of adults with no external incentive to exaggerate or to minimize their cognitive impairment. Although they might not be conspicuous, significant incentives to exaggerate impairment usually exist, even when the testing is conducted for purely clinical or research purposes. One reason is that most insurance companies which provide disability payments, including WCB, have a legal right to access reports of the results of any assessment performed on their client. This applies irrespective of the context or original purpose of the assessment. Hence, in practice, the results of *any* assessment could be used to adjudicate a claim and thereby impact a person's access to disability benefits. In support of this argument, the effort test failure rate was much higher in a group of fibromyalgia cases with disability claims than in those with no disability claim, even though the participants were told that the results were for research purposes and would not be entered into their file [10]. It is a misnomer to describe clinical assessments of people collecting disability payments as having no associated external incentive. The fact that the results could be used by an insurance company might explain relatively high failure rates on various effort tests in groups where it was assumed that there were no obvious external incentives because the assessments were done 'in a clinical setting' [11].

With these considerations in mind, the people we chose to comprise the third group were a series of children consecutively referred mainly by Social Services to the first author for neuropsychological assessment. Their diagnoses included a wide range of childhood disorders, including foetal alcohol spectrum disorder (FASD), autism, schizophrenia, ADHD, conduct disorder, Asperger's syndrome, mental retardation, learning disabilities and a variety of neurological disorders. It will be shown below that, in the absence of an external incentive to appear impaired, WMT effort test failure was rare, even when the person's cognitive deficits were objectively severe.

Method

Effort testing

The WMT is a test of verbal learning and memory which also allows evaluation of a person's effort to do well while taking the test, so a determination can be made about whether or not the person's test scores are valid estimates of ability [1, 4, 12]. All participants were given the computerized WMT as part of a complete neuropsychological or psychological assessment. In advance of the testing, all were informed that full effort was very important to ensure that valid test results would be obtained. The WMT measures the ability to learn a list of 20 word pairs (e.g., rat-tail) presented on a computer screen. The effort components of the WMT were designed to avoid confusing actual impairment with deliberate exaggeration. They are meant to be virtually insensitive to all but the most extreme forms of impairment of learning and memory and the range of genuine scores is very narrow. After being shown 20 word pairs on the computer screen, the person is required to choose the word from the original list in each of 40 new word pairs (e.g., 'rat' from 'rat-shoe'). This is the immediate recognition trial (IR), the first measure of effort. After a half-hour delay, the delayed recognition trial (DR) is presented, which is very similar to IR but it includes different foil words (e.g., 'rat-sock'). The patient is then given four separate measures of *memory* ability: the Multiple Choice subtest (MC), in which the person is shown the first word from each pair and is asked to choose the matching word from eight options; Paired Associates (PA), in which the person is given the first word from each pair by the tester and is asked to tell the tester the second word and Free Recall (FR), in which the person is asked to recall as many words as possible from the list in any order.

Tables and charts show the patient's scores compared with the mean scores obtained from any of 61 comparison groups in the WMT Windows program (Green, 2003), including patients with moderate to severe brain injury, epilepsy, fibromyalgia, mixed neurological patients, mentally retarded adults, and many other diagnostic groups. A cardinal feature of the WMT is that it consists of multiple subtests that vary widely in their objective difficulty level. Given the structure of the WMT, it is very difficult for a person who is not making a full effort to produce plausible scores on all subtests. In several independent studies, the WMT was 99–100% accurate in discriminating between people asked to try their best and those who were asked to simulate memory impairment [13–15]. Failure on the WMT in the latter studies and the current study was defined according to the standard criteria outlined in the test manual [1], based on scoring at or below given levels on IR, DR or the consistency between the two. The value of the cut-offs are not stated here to protect the integrity of the test.

Participants

There were three patient groups in our study. The first group consisted of individuals with an external incentive to appear impaired. They were 774 consecutively referred TBI cases who were referred for neuropsychological assessment to the second author. Within the TBI group, 577 cases could be classified as mild or complicated-mild TBI. They had a median post-traumatic amnesia (PTA) of zero and less than two days PTA in all cases. Their median Glasgow Coma Scale (GCS) score was 15 (mean = 14.6, SD = 1.1). The mean age of the mild TBI sample was 40.4 years (SD = 12) and 72% were male. They had a mean of 12 years of education (SD = 2.7) and were tested a median of 16 months post injury. Also within the TBI group, 197 cases could be classified as moderate-severe TBI. They all had a PTA of 24 hours or more (median 240 hours) and a median GCS score of eight (mean = 8.7, SD = 4). The mean age of the moderate-severe TBI sample was 39 years (SD = 13), 84% were male, they had a mean of 12 years of education (SD = 2.4) and they were tested a median of 19 months post-injury.

The results from CT or MRI brain scans were available from 177 of the latter group and brain abnormalities were present in 91% of these cases. In contrast, brain abnormalities, including extra-dural haematomas, were only noted in 24% of the 388 mild TBI cases, who had CT or MRI scans of the brain. 98% of the TBI cases were referred in the context of WCB claims (63% of cases), personal injury litigation (22% of cases) or disability insurance (13% of cases). These pending claims or existing benefits represented external incentives to appear impaired. Evaluations consisted of testing and interviewing for up to two days.

The second group was comprised of adults with a strong incentive to appear intact on neuropsychological testing. They were 118 parents seeking custody of their children, who were consecutively referred via Court order to the first author by the Department of Social Services for evaluation of their fitness to be parents. At the time of the assessment, their children were temporarily in the custody of the Department of Social Services. The mean age of the men was 36 years (SD = 8, $n = 31$) and of the women 32 years (SD = 8, $n = 85$). Both genders had an average of 11 years of education (SD = 2). In each case, all parents were interviewed, given personality testing and intelligence testing (Wechsler Abbreviated Scale of Intelligence (WASI) [16] and a documentary review was conducted. They were given the computerized WMT to measure effort and verbal memory [1]. Many parents were also given a problem solving test measuring abstract abilities [17].

A decision was given in each case reflecting the assessor's clinical judgment of whether the person was either fit or unfit to be a parent. The guidelines informing the decisions about fitness to be parents were taken from risk factors defined by Kuehnle, Coulter & Firestone [18]. In determining fitness to be parents, many factors were considered, apart from levels of intelligence or cognitive impairment. They included, among others, personality disorder, propensity for or history of child sexual, emotional or physical abuse or neglect, drug addiction, alcoholism, antisocial behavior, expressed tendencies towards sadism on personality testing and incapacity to empathize with others. In 55% of cases, the parents were judged unfit to take care of their children and in 10% a decision could not be reached, largely because some parents were highly defensive and unwilling to provide information. In only 35% of cases were the parents deemed fit to care for their children, indicating that, as a whole, this was a very low functioning group.

The third and final group consisted of individuals without an obvious external incentive to appear either impaired or to do well on testing. This group contained 258 children, aged seven to 17 years (mean age = 12.5; SD = 2.6) consecutively referred to the first author for neuropsychological assessment. There were 52 children with FASD, 54 with conduct disorder, 46 with learning disabilities, 41 with ADHD, 18 with schizophrenia or bipolar disorder, 14 with a primary diagnosis of autism or mental retardation and 33 children with miscellaneous diagnoses. The mean full scale intelligence of the child sample was 88.5 (SD = 14). Twenty-nine children had a mean verbal IQ of 70 or lower (mean = 64, SD = 5). There were a few cases with external incentives to appear impaired, such as two boys facing charges of murder and arson. However, in most cases, external incentives to appear impaired were absent.

Hypotheses

Effort to do well is thought to be the primary mediating variable affecting performance on the WMT effort subtests [1, 6, 12]. If so, we would expect to find the most failures on WMT in the group with an external financial incentive to appear impaired and the least failures in the parent group with an incentive to appear cognitively intact. The WMT failure rate in the child clinical group would be expected to be intermediate between those of the latter two groups.

If, on the other hand, WMT effort subtests were sensitive to true differences in cognitive abilities, we would expect to find (a) more effort test failures in the group with moderate to severe TBI than in those with relatively mild TBI; (b) more WMT failures in those with lower intelligence than in those of higher intelligence; (c) more WMT failures in the children than in adults with mild TBI and (d) more WMT failures in the parents seeking custody of their children than in adults with mild TBI.

Results

As predicted, the presence of external incentives was strongly related to rates of failure on WMT in the groups in this study. Table I shows the failure rates on WMT in the four main groups in this study.

The failure rate on the WMT in the parents seeking custody of their children (1.7% or 2/118) was markedly lower than the failure rate in both the more severe TBI group ($\chi^2 = 23$, d$f = 1$, $p < 0.001$) and in the mild TBI group ($\chi^2 = 65$, d$f = 1$, $p < 0.001$). Thus, the group with the greatest incentive to appear well on testing (parents) had the lowest failure rate on the WMT. In the group with the *least* severe brain injuries, the WMT failure rate was 40%, whereas it was only 21% in the group with the most severe brain injuries. This is a highly significant difference statistically ($\chi^2 = 23$, d$f = 1$, $p < 0.001$).

There were significant differences in FSIQ between the latter three groups ($F = 9.1$, d$f = 2$, 712, $p < 0.05$) but the differences are in the wrong direction to explain why so few parents failed the WMT. The mean FSIQ in the parents seeking custody of their children (93, SD = 15.7) was actually significantly lower than the mean FSIQ of the mild TBI group (99, SD = 14, $p < 0.001$ on Bonferroni post hoc comparisons). Thus, despite being less intelligent, the parents failed the WMT much less often than those in the mild TBI group. The parents did not differ significantly from the moderate-severe TBI group in FSIQ (mean = 96.5, SD = 13.5).

It may be pointed out that the validity of the FSIQ scores in anyone who fails the WMT effort subtests is under suspicion, unless the person suffers from dementia. The mean FSIQ was 104 (SD = 12) in the mild TBI cases who passed the WMT but it was only 91 (SD = 14) in those failing the WMT ($F = 99.7$, d$f = 1$, 440), indicating the magnitude of the effect of poor effort on IQ scores in this sample. The mean FSIQ was 98 (SD = 13) in the moderate-severe TBI

Table I. Rates of failure on WMT effort tests in groups with differing incentives.

Group	External incentive	Direction of incentive	n	Percent of group failing WMT*
Parents	Obtain child custody	Positive	118	<2%
Children tested clinically	None in most cases	Nil	259	11%
Moderate-severe TBI	Financial compensation	Negative	197	21%
Mild TBI	Financial compensation	Negative	577	40%

*Percentage scores rounded to nearest integer.

cases who passed the WMT but it was 90 (SD = 13) in those failing the WMT ($F = 9.4$, $df = 1, 159$). It was 93 (SD = 15) in the parents passing the WMT but 80 (SD 15) in the two parents who failed the WMT. Further evidence that low FSIQ does not account for WMT failure comes from the zero failure rate in the parents with a verbal intelligence score below 75 (range 52–74), as shown in Table II.

The two parents who failed the WMT were among the 62 parents deemed unfit to be parents. When these two women were questioned, it emerged that they changed their minds in the course of extended custody proceedings and did not actually want their children back at the time of testing. One was currently an inmate at a local prison, serving time for cocaine trafficking. Her WMT results strongly suggested poor effort because she scored higher on the harder of two subtests than on a much easier subtest. Her PA score was zero, meaning that in no case did she give the correct second word from a pair when prompted with the first word and this is extremely rare even in people with dementia. Nevertheless her score on the much harder FR test was within the normal range for adults of average intelligence (40%). This is a gross discrepancy, indicating invalid test results in this woman. The scores from the second parent failing the WMT, at first glance, might be regarded as a false positive classification for poor effort. However, she returned for testing a year later, by which time she was genuinely seeking custody of her children, and she easily passed the WMT on this occasion. Therefore, there would appear to have been no false positives in the parent custody sample, because only two out of 118 cases failed the WMT and they both admitted that they did not want their children.

Table II includes current findings, as well as data from several independently published studies, in which rates of failure on WMT were reported for groups with external incentives to do poorly or to do well on testing or with no such incentives. Table II shows that WMT failure rates are strongly associated with the presence of external incentives. If failure on the WMT effort subtests were affected by true differences in cognitive ability, we would expect to find significantly more WMT failures in the child group tested clinically than in adults with mild TBI. However, the failure rate on WMT was only 10.9%

Table II. WMT failure rates by external incentives.

Predominant external incentive	Group	*n*	Cognitive impairment expected	% failing WMT effort tests	Average WMT failure rate per group
Incentive to appear intact to gain child custody	Parents seeking child custody (VIQ < 75)	15	Yes	0%	1%
	Parents seeking child custody (VIQ ≥ 75)	103	No	2%	
No external incentive in most cases	Institutionalized mentally retarded adults (FSIQ < 70) Brockhaus & Merten, 2004	32	Yes	3%	10%
	Children tested clinically (VIQ < 75)	45	Yes	18%	
	Children tested clinically (VIQ > 75)	213	Yes	9%	
	Children tested clinically (FASD)	52	Yes	13%	
	Children tested clinically (ADHD)	41	Yes	12%	
	Adults with fibromyalgia (no disability claim and no benefits) Gervais et al., 2001.	50	No	4%	
Incentive to appear impaired to gain new disability benefits or compensation or to maintain existing benefits	Adults with fibromyalgia (already receiving or claiming disability benefits)* Gervais et al., 2001.	50	No	35%	36%
	Moderate-Severe TBI seeking disability/WCB/personal injury compensation	197	Yes	21%	
	Mild TBI seeking disability/WCB/personal injury compensation	577	No	40%	
	University students in ADHD assessment (Sullivan, in press) [7]	28	No	48%	

*This group had incentives to obtain or maintain disability benefits by appearing impaired but they were told that the test results were for research purposes and would not go on their file or affect their claim.

Table III. WMT scores by levels of full scale intelligence (FSIQ) in parents seeking custody.

FSIQ range	Mean FSIQ	WMT effort (mean of IR + DR)	WMT multiple choice	WMT paired assoc. recall	WMT free recall	Cat. test errors
(1) 49–82 ($n=29$)	73 (SD 7)	96% (SD 5)	83% (SD 17)	74% (SD 22)	48% (SD 16)	83 (SD 25)
(2) 83–102 ($n=52$)	92 (SD 5)	98% (SD 3)	91% (SD 12)	86% (SD 14)	56% (SD 15)	70 (SD 30)
(3) 102–126 ($n=31$)	112 (SD 7)	98% (SD 2)	97% (SD 6)	96% (SD 7)	70% (SD 13)	52 (SD 29)
Difference between groups 1 and 3	39 FSIQ points	2%	14%	22%	22%	−31 errors
Significance test	f 283, df = 2, 109	$\chi^2 = 3.8$, df = 2	$\chi^2 = 13.5$, df = 2	$\chi^2 = 24$, df = 2	f 13.5, df = 2, 109	f 7.3, df = 2, 97
Significance	$p < 0.0001$	$p < 0.02$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

Table IV. Failure rates on WMT in groups with mild or moderate-severe TBI and with or without intracranial abnormality on CT or MRI brain scan.

		CT/MRI			
		Normal		Abnormal	
Group	PTA duration	n	% failing WMT	n	% failing WMT
Mild TBI	Less than 24 hours	293	48%	94	33%
Moderate-severe TBI	24 hours or more	15	13%	155	23%

(28/258 cases) in the child group tested clinically, which is much lower than the 40% failure rate in those with mild TBI ($\chi^2 = 71$, df = 1, $p < 0.001$).

It is notable that the adult university students seeking ADHD assessments (Table II) failed the WMT in 48% of cases, whereas only 12% of children with ADHD in the current sample failed the WMT. This cannot be explained on the basis of lower cognitive abilities in the university students. Instead, it points towards poor effort as the reason for failure on the WMT. Within the child group, age did not significantly influence the failure rate on WMT ($\chi^2 = 7.6$, df = 10, $p = 0.65$), even though we know that children aged seven years will score lower on almost any ability test than children aged 17 years. Those children who failed WMT had a lower full scale intelligence (FSIQ = 81, SD = 13) than those who passed WMT (90, SD = 14; $f = 10.2$, df = 1, 256, $p < 0.01$) but, unlike age, FSIQ can be biased by poor effort. The FSIQ score reflects not only the child's intrinsic intelligence but also the amount of effort applied to the intelligence test. Ideally, if we wish to see how WMT scores are affected by intelligence, we would need a sample of people with high motivation to do well, so that effort is constantly good, while FSIQ varies.

Fortunately, we have an almost ideal group for this purpose in the parents seeking custody, whose FSIQ ranged from 49–126. If failure on the WMT effort subtests were affected by true differences in intelligence, we would expect to find significantly more WMT failures in the parents with the lowest intelligence compared with those in the highest range of intelligence. A comparison was made between three subgroups of parents in terms of their FSIQ (Table III). These subgroups differed significantly in terms of FSIQ ($p < 0.0001$), in terms of their error scores on the Category Test ($p < 0.001$) and in terms of their verbal memory, as measured by WMT FR ($p < 0.001$). The magnitude of the latter differences would be clinically important, such as the FSIQ difference of 39 points between those with a mean FSIQ of 73 vs. 112. However, there was no significant difference in the number of WMT failures in the parents across the three ranges of FSIQ shown in Table III (means 73, 92 & 112). No WMT failures were observed in the parents in the highest FSIQ range and only one failure occurred in each of the other two ranges.

The mean of the WMT IR and DR effort scores was 96% correct in the parents in the lowest range of intelligence (mean FSIQ = 73) and it was only 2% higher (98%) in the highest FSIQ group (mean FSIQ = 112). Such a difference is very small. Although statistically significant, a 2% difference between WMT mean scores is clinically insignificant. Both mean scores represent almost perfect accuracy and they are far above the cut-off, at which poor effort would be indicated.

Paradoxically, the failure rate on WMT in those TBI cases with a normal brain scan was significantly greater than the failure rate in those with an abnormal brain scan (Table IV, $Z = -4.8$, $p < 0.001$).

The effect of poor effort on neuropsychological test scores in general in the TBI group may be illustrated with reference to scores on the widely used Trail Making Test, Part B [19] in which the

Table V. Trail Making B scores by level of severity of brain injury and by pass or fail WMT.

Group	n	Mean trail making B (seconds)	SD	SE	95% Confidence interval for mean	
					Lower bound	Upper bound
Mild TBI	486	97	80	3.6	90	104
Moderate-severe TBI	170	101	72	5.5	90	112
Significance		Not Sig.				
Pass WMT	430	78	47	2.3	74	83
Fail WMT	226	136	105	7.0	122	150
Significance		$p < 0.001$				

person is asked to use a pencil to join together numbers and letters in alternating sequence (e.g., 1-A-2-B-3-C etc.). The score is the time taken to complete the task. Table V shows that there was no significant difference between 486 mild TBI and 170 moderate-severe TBI cases on Trail Making B. However, there was a highly significant difference between those who passed and those who failed the WMT effort subtests. We are justified in regarding the test scores as being unreliable in those failing the WMT. When we examine only the data from those who passed the WMT, a highly significant difference emerges. As we would expect, the 295 mild TBI cases performed significantly better on Trail Making B (mean = 71 seconds, SD = 36) than the 135 moderate-severe TBI cases (mean Trail Making $B = 92$ s, SD = 62, $f = 18$, d$f = 1$, 428, $p < 0.001$).

Discussion

In this study, we investigated external incentive as a determinant of effort on neuropsychological testing. We had a group with a strong financial incentive to appear impaired on neuropsychological testing, a group with a strong incentive to appear intact, and a group without a strong incentive to appear any particular way on testing. If, as we propose, external incentives impact performance on WMT effort tests more than true abilities, we would expect the group with an incentive to appear well to perform the best, the group without a strong incentive either way to appear in the middle and, finally, the group with an incentive to appear impaired to perform the worst and have the most failures on the WMT effort subtests. This is the exact trend which we found (Tables I and II).

Studies have shown that moderate to severe brain injury can have permanent effects on neuropsychological functioning, whereas mild TBI typically has no permanent effect or significantly lower and far less frequent permanent effects [20]. Hence, if WMT effort subtests (i.e., IR & DR) were affected by impairment in neuropsychological functioning, we would expect to see worse WMT performance in the moderate-severe TBI group than in the mild TBI group. However, in the current mild TBI group, the WMT was failed by almost twice as many as in those with moderate to severe TBI (40% vs. 21%). Additionally, we would expect to see more impaired performance and more failures in the group with brain abnormalities documented via CT or MRI scans than in the group with no abnormalities on brain scans. Yet the failure rate on the WMT was significantly higher in mild TBI cases with normal brain scans than in those with intracranial abnormalities on CT or MRI of the brain (Table IV). These data offer strong evidence that the WMT IR & DR effort subtests are not impacted by neuropsychological ability or brain lesions visible on imaging. The only group previously shown to perform below cut-offs when giving adequate effort to do well were those with dementia and verifiable clinical correlates of severe disability. Even moderate to severely brain injured individuals are still able to perform like healthy adults on WMT IR and DR subtests. Something other than true cognitive abilities is impacting performance on these subtests.

The powerful influence exerted by external incentives in determining performance on the WMT effort subtests was evident in the group of parents seeking custody of their children. They were motivated to appear unimpaired and they exhibited an extremely low WMT failure rate of only 1.7% (two failures out of 118 cases). Such a failure rate is much lower than in either group with TBI, even though many of the parents were of very low intelligence and had significant impairment in problem solving abilities (Table III). They were functioning at such low levels that 55% of them were judged to be unfit to regain custody of their children after the assessment. The parents in the lowest intelligence range had limited verbal memory capacity, as shown by definitely lower scores on FR of the WMT word list compared with the highest intelligence group. However, they still scored a mean of 96% correct on the WMT effort subtests (IR & DR) and there was no significant difference in WMT failure rates between those in the top or bottom ranges of intelligence (mean FSIQ 73 vs. 112, as in Table III). This demonstrates that the WMT IR and DR subtests are not sensitive to actual differences in intelligence or verbal memory, which do affect scores on WMT FR. It also shows that external incentives have a powerful influence on how people present during neuropsychological testing and it underscores the importance of data validity checks.

Extremely low IQ scores did not lead to WMT effort subtest failures. In the group of parents seeking custody of their children who had verbal IQs below 75, including a woman with a FSIQ of 49, not a single individual failed the WMT effort subtests (Table II). This is consistent with previous research demonstrating that mild mental retardation, by itself, does not impact performance on the WMT effort subtests. Brockhaus & Merten [15] reported that 31 out of 32 institutionalized but testable mentally handicapped adults in Germany were able to pass the WMT using the usual criteria (Table II), showing that low cognitive abilities, in themselves, are not necessarily associated with failure on the WMT. On the contrary, the single group in Table II with the highest levels of education and intelligence were university students, yet they had the highest WMT failure rate of all. Sullivan [7] reports a 48% failure rate on the WMT in university students seeking benefits and special accommodation for claimed ADHD. In contrast, the WMT failure rate was only 12% in ADHD children in the current study, who were on average, only 10.9 years old (Table II). Harrison [21] has also written about incentives for symptom exaggeration in adult students seeking classification as having learning disabilities or ADHD. Taken together, these studies show how external incentives to appear a certain way, either intact or impaired, have a strong moderating influence on presentation during cognitive and psychological evaluations.

In the children with serious psychiatric and neurological conditions, such as FASD, there were few external incentives to perform well or to do poorly. The WMT was passed by 89% of these 258 children. Their failure rate on WMT was significantly lower than in the mild or moderate to severe TBI groups. There was no age effect and WMT failure rates were spread equally across the range seven to 17 years. It is likely that, in the few children who did fail the WMT (10.9% of cases), poor effort was the underlying cause in most cases. It was demonstrated in a previous study that the few children who failed the WMT on first testing could easily be induced to pass the WMT on retesting by being given a small external incentive in the form of a can of pop or a candy [8], once again highlighting the importance of motivation on testing.

There are striking counterintuitive differences in WMT failure rates between groups in the current study and in published comparison groups. The WMT failure rates were, in rank order, 0% in parents with a verbal intelligence less than 75; 3% in mentally retarded adults in an institution; 3% in adults declared unfit to be parents; 10.9% in children tested clinically; 21% in adults with moderate to severe TBI; 40% in adults with mild TBI and 48% in university students seeking ADHD assessment. These differences cannot be explained on the basis of intelligence, years of education, age or any actual cognitive deficits. They can only be explained by the fact that some had incentives to appear impaired, whereas others had an incentive to look good.

The monetary incentive explains why almost twice as many cases of mild TBI failed the WMT compared with moderate to severe TBI. Most cases of moderate to severe brain injury within the WCB system would initially be given full benefits automatically and only later would there be incentives to perform poorly on testing in an attempt to maximize the duration or amount of benefits paid. Most cases would be sent for a neuropsychological assessment. In many cases of minor head injury, on the other hand, especially those with normal brain scan results, benefits would be granted less generously and terminated much sooner, thereby providing incentives for the mild TBI cases to exaggerate more often than those with more severe injuries. Those people with mild TBI who did not complain and who returned to work promptly would probably not be referred for neuropsychological assessment. Hence, there were unintended selection biases contributing to the very high failure rate in the current mild TBI sample. It is probable that an unselected series of all consecutive cases of mild TBI admitted to a hospital would yield much lower WMT failure rates.

The current finding of a major excess of WMT failure in people with mild TBI compared with moderate-severe TBI contrasts with the report of Bowden, Shores & Mathias [22], in which there was no difference in WMT failure rates between mild and severe TBI in a mixed sample of 100 children and adults. The difference may be partly explained by the relatively small sample in the Bowden et al. study [22], with especially small numbers in the severe category, and the fact that children as young as six years old were tested with the WMT. The WMT is not recommended for children under the age of seven years nor for anyone with a reading level of less than grade three [8]. It was not stated in the paper how many in the sample were adults and how many were children but we find much lower rates of failure on WMT in children than in adults.

Hence, the relationship between WMT failure and severity of brain injury needs to be studied in adults and children separately. The authors did not use the WMT in the standard way in which it was designed to be used and as described in the test manual. They used only one subtest out of six, the IR subtest. To define failure on the WMT, at least the first two subtests plus a consistency score are needed in the standard administration procedures.

We may nevertheless test Bowden et al.'s [22] main hypothesis, which was that 'all of the scores from the WMT measure memory or some component of general cognitive ability' (p. 868, para. 2). It should be pointed out that, if the WMT effort subtests really did measure memory and general cognitive ability, we would expect to find significantly *lower* WMT effort scores in their severe TBI than in their mild TBI groups, whereas they actually reported no difference between these groups in their own study. The absence of any difference between the mild and severe TBI cases can easily be explained by equal amounts of effort being applied in that small sample by the mild vs. the severe TBI cases or possibly by inequalities in the numbers of adults and children in their sample. However, their finding of an absence of any advantage on WMT in the mild vs. the severe TBI group cannot be explained by the hypothesis that the WMT is sensitive to levels of ability. Thus, the data from the Bowden et al. [22] study actually support the insensitivity of the WMT effort measures to true differences in memory ability. Other arguments against the latter hypothesis follow:

(1) We must assume that mentally retarded adults, who are institutionalized and incapable of living independently, have lower cognitive abilities than university students. If Bowden et al.'s [22] hypothesis were correct, we would find more WMT failures in the mentally handicapped than in university students. Instead we find the opposite, which is only a 3% failure rate on the WMT in mentally retarded adults but a 48% failure rate in university students seeking ADHD assessment (Table II).
(2) If Bowden et al.'s [22] hypothesis were true, we would find fewer WMT failures in university students undergoing ADHD assessments than in children with ADHD. However, on the contrary, we actually find the university students of Sullivan [7] failing the WMT four times more often than children with ADHD (Table II).
(3) If the Bowden et al.'s [22] hypothesis were correct, we would confidently predict more WMT failures in the moderate-severe TBI group of the current study than in the mild TBI group but this was not the case. The mild TBI group actually failed WMT roughly twice as often as those with more severe brain injuries. This is opposite to what Bowden et al. [22] would have predicted.
(4) If their hypothesis were true, we would find fewer WMT failures in university students undergoing ADHD assessments than in children or adults with a verbal IQ below 75. On the contrary, we actually find that the university students of Sullivan [7] failed the WMT far more often than either adults or children of very low verbal intelligence (Table II).
(5) We would be hard pressed to reconcile the current findings in people with mild TBI with Bowden et al.'s hypothesis [22] that the WMT effort subtests measure actual cognitive abilities. If so, why do 40% of those with mild TBI fail the WMT, whereas 98.3% of parents undergoing custody assessments pass the WMT, including a woman with a FSIQ of 49?
(6) Why do we find such a high rate of failure on extremely easy WMT subtests in the mild TBI group, whereas a vast literature on sports concussions and in mild TBI generally shows no measurable neuropsychological deficits a week after loss of consciousness? [20].
(7) Only 3% of mentally handicapped adults failed the WMT [15]. Does mild TBI really produce cognitive deficits markedly greater than those seen in mentally handicapped adults? Is that why people with mild TBI fail WMT IR and DR over ten times more often than mentally handicapped adults and several times more often than children with disabling conditions like FASD?

Clearly, in the current study, the WMT effort subtests do not behave like memory tests or tests of general cognitive ability. The results on WMT IR and DR in Tables I, II and III cannot be explained in terms of actual differences in ability levels between groups. The findings of this study are consistent with previous research in people with fibromyalgia, demonstrating that applying for disability payments or currently receiving them represents a strong external incentive that impacts performance on the WMT subtests far more than cognitive abilities or symptom severity [10]. The failure rate on the WMT effort subtests in the latter study was 35% in the disability claimants but it was only 4% in the non-disability cases (Table II).

In the current study, for a contrast, we also had a group with a powerful positive incentive to appear unimpaired. Yet even in the lowest FSIQ range (mean FSIQ = 73), less than 2% of the parents failed the WMT. Two parents with a VIQ above 75 scored below the WMT effort subtest cut-off points. They might be construed as false positives if they were motivated to do well and failed the WMT. However, the two cases who scored below cut-offs were, in fact, not trying to obtain custody of their children. It can be inferred that they were not trying their best on testing because they did not want to be declared fit to be parents. Indeed, one of the parents who later changed her mind and did sincerely want to regain custody was able to pass the WMT effort subtests with ease at a later testing date.

Whereas mild TBI results in minimal or no permanent measurable cognitive deficits [20], those in the mild TBI group failed the WMT twenty-three (23) times more often than the parents seeking custody. This is a significant fact for anyone trying to assess the level of actual cognitive impairment present in people with mild TBI, who are receiving disability from any source or who are trying to obtain compensation for their injuries. It implies that, in a large proportion of such cases, there will be an artificial suppression of scores on cognitive tests, as a function of poor effort and/or cognitive symptom exaggeration. In group data, the excess of poor effort in mild vs. moderate-severe TBI can make it seem that the cognitive deficits are equivalent in these two groups (Table V). The false impression arises from the suppression of scores in those making a poor effort, resulting in their test scores being invalid. In turn, this obscures real group differences between mild and moderate to severe TBI in cognitive abilities. If effort is not controlled, as we see in Table V, the mild and moderate-severe TBI groups score the same as each other on Trail Making B. However, when the poor effort cases are dropped from the analysis there is a clearly significant difference in the expected direction. Moderate-severe TBI does cause significantly lower cognitive ability than mild TBI but this important fact is often obscured by the effects of poor effort.

If we failed to measure effort and if we assumed that all neuropsychological test data were valid in the TBI cases of the current study, we would wrongly conclude that severe cognitive deficits were present in many mild cases of TBI. The WMT helps to identify those in whom effort is not sufficient and whose test scores underestimate their actual ability. It is advisable to use well validated effort tests like WMT whenever any cognitive assessment is performed in people with TBI [3], especially in the majority of cases in whom the TBI is relatively mild.

One limitation to the present study is the absence of cases with extreme degrees of cognitive impairment, such as may be found in people with advanced dementia. Notwithstanding the fact that a woman with a FSIQ of 49 passed the WMT in this study, people with dementia sometimes fail the WMT. Additional studies investigating the WMT effort subtests in individuals with dementia to determine base rates and performance profiles would be helpful. They are especially interesting for studying ways of identifying false positives on the WMT (i.e., WMT failure despite best effort). The WMT program already contains some methods to assist in identifying such cases, as well as including data from groups with dementia [1].

The primary goal of this study was to investigate performance on the WMT effort subtests in people with a wide range of age, intelligence and cognitive abilities and with varying degrees of external incentives to perform well or to perform poorly on neuropsychological and psychological testing. The main question was how much influence incentives would exert on WMT effort subtest failure, as opposed to differences in actual cognitive abilities. To paraphrase the title of a widely cited paper on recovery from head injury [23] the results show that ‘incentive matters’. Major measures of general cognitive abilities, such as intelligence test scores, do not account for WMT failure. The results show very clearly that, in the groups studied, it is the presence of external incentives to appear impaired which is the primary predictor of WMT effort test failure.

Acknowledgement

Thanks to Dr Laura L. S. Howe for her keen editing work.

References

1. Green P. Manual for the Word Memory Test for Windows. Edmonton: Green’s Publishing; 2003.
2. Bush S, Ruff R, Troster A, Barth J, Koffler S, Pliskin N, Reynolds C, Silver C. Symptom validity assessment: Practice issues and medical necessity (NAN Policy and Planning Committee). Archives of Clinical Neuropsychology 2005;20:419–426.
3. Iverson G. Ethical issues associated with the assessment of exaggeration, poor effort and malingering. Applied Neuropsychology 2006;13:77–90.
4. Green P, Astner K. Manual for the oral Word Memory Test. CogniSyst, NC;1995.
5. Ardolf BR, Denney RL, Houston CM. (in press). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. The Clinical Neuropsychologist.
6. Green P, Rohling ML, Lees-Haley PR, Allen L. Effort has a greater effect on test scores than severe brain injury in compensation claimants. Brain Injury 2001;15:1045–1060.
7. Sullivan B, May K, Galbally L. Symptom exaggeration by college adults in ADHD and learning disorders assessments. Applied Neuropsychology (in press).
8. Green P, Flaro L. Word Memory Test performance in children. Child Neuropsychology 2003;9:189–207.

9. Larrabee GJ. Assessment of malingering. In: Larrabee GJ, editor. Forensic Neuropsychology: A scientific approach. New York: Oxford University Press; 2005. pp 115–158.
10. Gervais RO, Russell AS, Green P, Ferrari R, Pieschl SD. Effort testing in patients with fibromyalgia and disability incentives. Journal of Rheumatology 2001;28:1892–1899.
11. Curtis K, Greve K, Bianchini K, Brennan A. California Verbal Learning Test indicators of malingered neurocognitive dysfunction. Assessment 2006;13:46–61.
12. Green P, Lees-Haley PR, Allen LM. The Word Memory Test and the validity of neuropsychological test scores. In: Hom J, Denney RL, editors. Detection of response bias in forensic neuropsychology. New York: Haworth Medical Press; 2003.
13. Iverson G, Green P, Gervais R. Using the Word Memory Test to detect biased responding in head injury litigation. The Journal of Cognitive Rehabilitation 1999;17:4–8.
14. Tan JE, Slick D, Strauss E, Hultsch D. How'd they do it? Malingering strategies on symptom validity tests. The Clinical Neuropsychologist 2002;16:495–505.
15. Brockhaus R, Merten T. Neuropsychologische Diagnostik suboptimalen Leistungsverhaltens mit dem Word Memory Test. Nervenartz 2004;75:882–887.
16. Zhu J. Wechsler Abbreviated Scale of Intelligence. San Antonio: The Psychological Corporation; 1999.
17. De Filippis N, PAR staff. The Category Test (computer program). Psychological Assessment Resources; 1979 & 2005.
18. Kuehnle M, Coulter J, Firestone G. Child protection evaluations: The forensic stepchild. Family and Conciliation. Courts Review 2000;38:368–391.
19. Reitan RM. Manual for administration of neuropsychological test battery: Theory and interpretation. Indianapolis, IN: Neuropsychological Press; 1969.
20. Iverson G. Outcome from mild traumatic brain injury. Current Opinion in Psychiatry 2005;18:301–317.
21. Harrison A. Adults faking ADHD: You must be kidding. ADHD report. New York: Guilford Press; 2006;14:1–6.
22. Bowden S, Shores A, Mathias J. Does effort suppress cognition after traumatic brain injury? A re-examination of the evidence for the Word Memory Test. The Clinical Neuropsychologist 2006;20:858–872.
23. Binder L, Rohling M. Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. American Journal of Psychiatry 1996;153:7–10.