2011

Assessment of Suboptimal Effort Using the CVLT-II Recognition Foils: A Known-Groups Comparison

Matias Mariani
University of Windsor

Recommended Citation

Mariani, Matias, "Assessment of Suboptimal Effort Using the CVLT-II Recognition Foils: A Known-Groups Comparison" (2011). *Electronic Theses and Dissertations*. 491.
https://scholar.uwindsor.ca/etd/491

Assessment of Suboptimal Effort Using the CVLT-II Recognition Foils:

A Known-Groups Comparison

by

Matias Mariani, M.A.

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada
2009

Author's Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

The present study sought to generate an embedded effort index within the CVLT-II yes/no recognition trial using a known-groups design. Four types of recognition foils—i.e., novel/semantically unrelated (UN), novel/semantically related (PR), list B/semantically unrelated (BN), and list B/semantically related (BS)—as well as two composites—i.e., easy to reject foils (ETR) and difficult to reject foils (DTR)—were evaluated on their ability to distinguish between a group of 82 outpatients with moderate-severe traumatic brain injuries (TBI) and a group of 31 litigants meeting Slick et al. (1999) criteria for malingered neurocognitive dysfunction (MND). Separate multiple logistic regression analyses were performed. The full model based on the 4 foils correctly classified 88.5% of cases (61.3% sensitivity/98.8% specificity). The full model based on the composites correctly classified 81.4% of cases (45.2% sensitivity/95.1% specificity). With respect to univariate predictors, UN correctly classified 51.6-64.5% of MND cases and 90.2-100% of TBI cases depending on the diagnostic cut-off used. ETR also showed good classification accuracy (25.8-51.6% sensitivity/90.2-100% specificity). Three different ratios were generated from the original analyses—UN/PR, UN/(PR+BN+BS), and ETR/DTR. All three ratios yielded good to excellent diagnostic accuracy (87% sensitivity/98.4% specificity, 70.4% sensitivity/97% specificity, and 38.5% sensitivity/95.5% specificity, respectively). In addition, UN, ETR, and the multivariate equations were cross-validated with a group of 19 patients with complicated mild TBI supplying adequate effort (MTBI) and a group of 23 patients with complicated mild TBI performing poorly on effort measures (SE), resulting in high specificity values depending on the cut-offs used. Finally, previous research using the CVLT and CVLT-II

(Coleman et al., 1998; Curtis et al., 2006; Millis et al., 1995; Millis et al., 2007; Sweet et al., 2000) was replicated.

Overall, the UN variable, the ETR composite, both multivariate equations, and all three ratios derived from the foils of the CVLT-II yes/no recognition trial show considerable merit as embedded effort indices. Positive and negative predictive power values are provided for all predictors at various diagnostic cut-offs across 5 hypothetical base rates in order to facilitate generalization of findings to different settings. Clinical and forensic implications are discussed with a focus on differential diagnoses.

Dedications

This manuscript is dedicated to my wife Melanie, my parents Aldo and Edith, and my siblings Marcos, Carolina, and Lucas. My accomplishments could not have been possible without your support, patience, understanding, and love.

Acknowledgements

The author gratefully acknowledges Dr. D. Shore for his mentorship and support throughout the author's graduate career.

In addition, the author is appreciative of Dr. R. Hanks for allowing access to her office resources and case files at the Research Institute of Michigan (Detroit, MI) and Psychological Systems Inc. (Royal Oak, MI); Dr. S. Millis for providing valuable statistical advice and comments; and Mr. R. Kotasek for assisting with the initial steps of database preparation.

Finally, the author would like to thank Dr. A. Baird, Dr. D. Jackson, Dr. P. Graniero, and Dr. B. Axelrod for their constructive feedback and recommendations throughout the different stages of the dissertation process, Dr. T. Snyder, Ms. J. Koushik, and Ms. M. Fielding for their comments on the manuscript, and Dr. N. Fisher for her guidance.

Table of Contents

List of Abbreviations

ACRM	American Congress of Rehabilitation Medicine
AF	Affective Disorders Scale, SIMS
AM	Amnestic Disorders Scale, SIMS
APA	American Psychiatric Association
BMA	Bayesian Model Averaging
BN	Semantically Unrelated List B Recognition Item, CVLT-II
BS	Semantically Related List B Recognition Item, CVLT-II
CIA	Critical Item Analysis, CVLT-II
CDC	Centre for Disease Control and Prevention
CVLT	California Verbal Learning Test, (-II) Second Edition
d’	Recognition Discriminability, CVLT and CVLT-II
DSM-IV TR	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision
DR	Delayed Recognition, WMT
DTR	Difficult to Reject Composite, CVLT-II
ETR	Easy to Reject Composite, CVLT-II
FBS	Fake Bad Scale, MMPI-II
FCR	Forced-choice Recognition Total, CVLT-II
FMS	Failure to Maintain Set, WCST
FTT	Finger Tapping Test, (-d) dominant hand, (-nd) non-dominant hand
GCS	Glasgow Coma Scale
IR	Immediate Recognition, WMT
JOLO	Judgment of Line Orientation
LDCR	Long-Delay Cued Recall, CVLT and CVLT-II
LDFR	Long-Delay Free Recall, CVLT and CVLT-II
LI	Low Intelligence Scale, SIMS
LOC	Loss of Consciousness
MMPI-II	Minnesota Multiphasic Personality Inventory, Second Edition
MND	Malingered Neurocognitive Dysfunction
NI	Neurological Impairment Scale, SIMS

NIM	Negative Impression Management Scale, PAI
NPP	Negative Predictive Power
P	Psychosis Scale, SIMS
PAI	Personality Assessment Inventory
PPP	Positive Predictive Power
PR	Prototypically Related Novel Recognition Item, CVLT-II
PTA	Post-traumatic Amnesia
PTSD	Post-traumatic Stress Disorder
RDS	Reliable Digit Span, WAIS-R and WAIS-III
RIM	Rehabilitation Institute of Michigan
RMT	Recognition Memory Test
ROC	Receiver Operating Characteristic
SDCR	Short-Delay Cued Recall, CVLT and CVLT-II
SDFR	Short-Delay Free Recall, CVLT and CVLT-II
SDMT	Symbol Digit Modality Test, (-O) Oral Version, (-W) Written Version
SEMTBIS	South-eastern Michigan Traumatic Brain Injury System
SIMS	Structured Inventory of Malingered Symptomatology
SR	Sentence Repetition
SVT	Symptom Validity Test
TBI	Traumatic Brain Injury
TMT	Trail Making Test, (-A) Part A, (-B) Part B
TT	Token Test
TOMM	Test of Memory Malingering
UN	Semantically Unrelated Novel Recognition Item, CVLT-II
V-DS	Vocabulary-Digit Span Discrepancy, WAIS-R and WAIS-III
VFD	Visual Form Discrimination
WAIS	Wechsler Adult Intelligence Scale, (-R) Revised, (-III) Third Edition
WCST	Wisconsin Card Sort Test
WMT	Word Memory Test

Introduction

A traumatic brain injury (TBI) is any damage to the brain caused by an external mechanical force applied to the head (South-eastern Michigan Traumatic Brain Injury System [SEMTBIS], 2004). TBIs can occur from acceleration or deceleration forces and/or physical deformation of the skull from blunt trauma to the head (i.e., closed head injury; e.g., via motor vehicle accidents, falls, sports collisions, etc.), which may subsequently cause diffuse axonal injury, focal axonal shearing, contusions, subdural haematomas, and intracerebral haemorrhages (Bigler, 2001; Greenberg, Aminoff, & Simon, 2002; Rao & Lyketsos, 2000). Alternatively, TBIs can result from penetrating objects (e.g., gunshot wounds, open skull fractures, etc.), which result in severe lacerations of brain tissue (Blumenfeld, 2002; Grubb & Coxe, 1974). Secondary damage from TBIs can result from cerebral oedema, hypoxia, ischemia, compromised cerebral vasculature, and increased intracranial pressure, which may consequently cause a herniation syndrome (Bigler, 2001; Blumenfeld, 2002). Due to the range of mechanisms of injury, the clinical presentation of TBI can vary. In addition, TBI appears highly susceptible to feigning and exaggeration because its symptoms are generally non-specific and diffuse, and because there is a wealth of information readily accessible to the public (Tan, Slick, Strauss, & Hultsch, 2002; Wise, Oliveira, Lacy, Han, & Pyykkonen, 2006). Thus, it is imperative for clinicians to conduct thorough assessments and consider all possible differential diagnoses when formulating cases presenting with possible head trauma.

Consequently, the aims of this paper are as follows. First, this paper reviews the clinical sequelae normally observed in patients with TBI and the variables most commonly used to stage severity of injury—namely, loss of consciousness (LOC), post-traumatic amnesia (PTA), Glasgow Coma Scale score (GCS; Teasdale & Jennett, 1974), and significant findings on neuroimaging. Second, this paper examines the definition of malingering, its prevalence in medical and forensic settings, its diagnostic criteria, and its differential diagnoses. Third, it highlights the stand-alone tests and embedded indices that are most frequently used to determine suboptimal effort output, symptom exaggeration, and negative response bias. Fourth, an emphasis is placed on evaluating the California Verbal Learning Test – second edition (CVLT-II; Delis, Kaplan, Kramer, & Ober, 2000) as a potential tool for detecting suboptimal effort output. More specifically, this study investigates the utility of the CVLT-II yes/no recognition foils to differentiate between patients with moderate-severe TBI and litigants with questionable head injuries putting forth suboptimal effort. From the yes/no recognition foils, it is hypothesized that those novel and semantically unrelated to the target list items will be the best predictors of suboptimal effort output in this known-groups design. Finally, the variables found to distinguish between moderate-severe TBI and litigants with insufficient effort are cross-validated with two additional samples—patients with complicated mild TBI supplying adequate effort and patients with complicated mild TBI performing poorly on effort measures—in an effort to generalize the findings. Collectively, the overarching goals of this study are to develop a new effort index embedded within the CVLT-II yes/no recognition trial that has high sensitivity and specificity and can help reduce time spent administering stand-alone symptom validity tests while supplementing clinical and forensic decision-making.

TBI

The paramount feature of TBI is a loss of consciousness (LOC) following trauma, with increased lengths of LOC being associated with poorer prognosis. In general, the symptoms that accompany a TBI may include headaches and neck pain, confusion and disorientation, dizziness, fatigue, sleep disturbances, memory problems, mood changes, slowed processing, and difficulties with attention and concentration, as well as nausea, blurred vision, tinnitus, hypersensitivity to stimuli, and loss of smell or taste (Centers for Disease Control and Prevention [CDC], 1999; Rao & Lyketsos, 2000; Ziino & Ponsford, 2005). With this extensive list of symptoms, differential diagnosis can be a daunting task. In addition, the nomenclature of injury severity in TBI is somewhat inconsistent, especially when dealing with cases at the mild end of the severity spectrum. Because of the ambiguity in symptomatology and inconsistency in classification of injury severity and prognosis, Teasdale and Jennett (1974) devised the Glasgow Coma Scale (GCS). The GCS is a widely used instrument that quantifies the depth and length of a coma based on 3 types of responses to external stimuli (i.e., eye, verbal, and motor) that are assessed as early as possible following trauma. Based on the patient’s score, the brain injury can be classified as *mild* (13-15), *moderate* (9-12) and *severe* (3-8) (Teasdale & Jennett, 1974). By formulating a standard for injury indexing, Teasdale and Jennett afforded professionals from different fields the ability to communicate information easily between each other when treating a comatose patient. However, because the GCS is time-dependent (i.e., it is administered during triage) and because the arrival of medical assistance varies widely, the time between the point of injury and test administration tends to vary, which may lead to some patients being classified as having less severe injuries than they may have (Ruff & Jurica, 1999). Moreover, because the symptoms of TBI may not surface immediately following injury (Reitan & Wolfson, 2000), some patients may be undiagnosed altogether if the GCS is used as the sole severity indexing tool. In addition, GCS scores tend to be poorer with the presence of alcohol intoxication (Jagger, Fife, Vernberg, & Jane, 1984), indicating that some patients may receive worse scores during the initial assessment but may show an improvement in neurocognitive status once their inebriation has subsided. Thus, although the GCS was a vast improvement towards the standardization of injury classification, the variability of contextual factors surrounding the time of injury makes its use somewhat limited with respect to atypical cases. Finally, because the GCS is meant to be used in the first few hours following injury (i.e., acute phase), its purpose is limited when used retroactively. These caveats notwithstanding, GCS is the most widely used tool in the literature for estimating severity of brain injury following trauma.

Alternatively, some clinicians rely on LOC and posttraumatic amnesia (PTA) to base their estimates of severity of injury because these variables can be used retrospectively. LOC is normally calculated by the time it takes for the patient to be able to follow commands such as “raise your hand” or “stick out your tongue.” PTA is defined as the period of time from the point of injury until the individual has continuous recall of ongoing events (Whyte & Rosenthal, 1988). Because PTA is assessed through direct querying, it can only be determined after LOC has subsided. Unfortunately, as with many aspects of assessment, there is a degree of clinical judgment involved in determining when responses are “continuous” enough for PTA to be considered lifted, which adds a degree of uncertainty to injury classification. Nonetheless, PTA has been reported as correlating well with GCS (Levin, Grossman, & Benton, 1982, as cited in Lezak, 1995) and as being a strong predictor of outcome at 18 months and 3 years post-injury (Tate, Harris, Cameron, Myles, Winstanley, Hodgkinson, Baguley, & Harradine, 2006). Overall, depending on the duration of a LOC and PTA as well as the presence of focal neurological signs, the prognosis of a TBI can vary from mild and transient to severe and permanent (Blumenfeld, 2002; Dikmen, Machamer, Winn, & Temkin, 1995; Reitan & Wolfson, 2000).

Bigler (1988) classifies individuals as having sustained a moderate TBI if they present with LOC lasting at least 1 hour and PTA lasting up to 24 hours. In the event that the individual is alert but displays focal neurological signs and PTA, a moderate TBI classification is also warranted. A TBI is considered severe when a patient is fully comatose for more than 1 day, has PTA lasting between 1 and 7 days or more, and shows motor deficits and pathologic reflexes (Bigler, 1988; Rao & Lyketsos, 2000). Recovery following a moderate TBI is favourable but not complete, whereas recovery following a severe TBI is more limited, particularly in older adults (Goleburn & Golden, 2001; Goldstein & Levin, 2001). Although there is some improvement expected to occur in cognitive functioning in patients with moderate-severe TBI within the first two years after injury, their cognitive profiles remain significantly impaired compared to controls beyond two years post-injury (Schretlen & Shapiro, 2003). In general, there appears to be a consensus as to what constitutes moderate and severe TBI. With respect to the classification of mild TBI, however, the inclusion criteria differ slightly depending on the stringency employed by specific studies as well as depending on the different settings (e.g., clinical, medico-legal, etc.).

Mild TBI, also known as a concussion, is defined as a "reversible impairment of neurological function for minutes to hours following a head injury" (Blumenfeld, 2002, p. 142). Bigler (1988) defines it as a transient loss or alteration in consciousness in the absence of definite localizing or lateralizing signs, and accompanied by an amnesic period lasting no longer than 1 hour. Additionally, although the text revision of the 4th edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric Association [APA], 2000) does not have a classification specifically for mild TBI (or any TBI except for "dementia due to head trauma," p. 164), the proposed diagnostic criteria for postconcussional disorder appear comparable to other mild TBI classifications. Specifically, the DSM-IV-TR criteria involve a transient LOC lasting no longer than 5 minutes, followed by PTA lasting less than 12 hours (APA, 2000). Conversely, other researchers define mild TBI as a period of LOC less than or equal to 20 minutes and/or accompanied by brief PTA, a GCS of 13 and above, no focal neurological deficit, no intracranial complications (e.g., haemorrhage), and normal findings on neuroimaging (Alexander, 1995; Gennarelli, 1986; Goldstein & Levin, 2001).

Finally, the American Congress of Rehabilitation Medicine (ACRM; 1993) also established guidelines to classify mild TBI in efforts to clear the diagnostic picture. The criteria used by the ACRM included at least one of the following characteristics: any LOC or alteration of mental state following trauma that is less than 30 minutes in length, any loss of memory immediately before or after the trauma lasting less than 24 hours, focal neurological deficit that may or may not be transient, and a GCS score of at least 13 as assessed 30 minutes post-trauma (ACRM, 1993). However, taking any of these criteria alone (e.g., LOC< 30 minutes) in the absence of other criteria would cause potential discrepancies in the classification systems. For example, if an individual presents with 20 minutes of LOC, they would not be considered to have mild TBI according to the DSM-IV-TR but they would still meet the mild TBI criteria as defined by Alexander (1995). In addition, a person displaying a PTA period lasting 22 hours would be considered to have a moderate TBI according to Bigler (1988) but a mild TBI according to the ACRM. Thus, there appears to be quite a bit of overlap between injury labels depending on the different classification systems. As a result, Dikmen, Machamer, and Temkin (2001) compared the different indexing cut-offs used by different researchers and clinicians in efforts to clear the ambiguity regarding the definition of mild TBI as well as to assess how well those specific injury severity variables correlate with neuropsychological tests sensitive to brain injury. Specifically, they divided a sample of individuals with mild TBI into four groups according to different measures of head injury severity and evaluated the neuropsychological profiles of each group against a group of trauma patients without head injuries. Group 1 consisted of individuals with GCS scores 13-15 but no other restrictions, whereas Groups 2 through 4 had additional criteria (LOC < 1 hour, negative CT scans, and PTA <= 24 hours) increasing in stringency whereby Group 4 individuals had to meet all the aforementioned criteria. The results showed that Group 1 differed from the control group on verbal recall, but on no other task at 1 month post-injury. Additionally, there were no other significant differences between the experimental groups and the control group on any measures at 1 month or 12 months post-injury. In general, the more stringent the criteria used to define mild TBI, the less likely there were any differences in neuropsychological performance compared to the control group at follow-up. Thus, a conservative approach for defining a mild TBI group would involve a GCS score of 13 or higher, LOC less than 1 hour, negative findings on neuroimaging, and PTA up to 24 hours. Groups meeting these criteria would be expected to have complete recovery of neuropsychological functioning beyond 1 month post-injury.

On somewhat rare occasions, individuals have been reported as experiencing prolonged effects following a mild TBI—known as persistent post-concussive syndrome (Alexander, 1995; Blumenfeld, 2002)—which may be caused and/or maintained by premorbid factors such as chronic emotional or psychological distress, neurological or psychiatric illness, learning disability, alcohol abuse, and tendency for somatization, as well as low pre-injury baseline functioning (Babin, 2002; Dikmen et al., 2001; Millis & Volinsky, 2001; Rao & Lyketsos, 2000; Reitan & Wolfson, 2000). Likewise, when an individual's measures of acute injury severity fall within the mild TBI range but there is evidence of neurological insult, the TBI is classified as "complicated" mild because its effect on neuropsychological sequelae is comparable to moderate TBI (Heinly, Greve, Bianchini, Love, &Brennan, 2005; Millis, Putnam, Ricker, & Adams, 1995).

However, with respect to *uncomplicated* mild TBI, complete recovery to premorbid levels of functioning is typically reached within 1 to 3 months post-trauma, with most rapid recovery occurring within the first few weeks post-injury (Dikmen et al., 1995; Gentilini, Nichelli, & Schoenhuber, 1989; Goldstein & Levin, 2001; Ponsford et al., 2000). As a rule of thumb, there is a dose-response relationship between injury severity and cognitive impairment (Rohling, Meyers, & Millis, 2003; Schretlen & Shapiro, 2003), whereby more impaired patients tend to perform worse on neuropsychological tests. Thus, whenever individuals report constant and pervasive symptoms beyond the 3-month time frame that are inconsistent with the severity of injury after all complications have been ruled out, a secondary explanation (e.g., psychological, motivational, etc.) must be entertained as the possible source of the symptomatology. Above all, the question of suboptimal effort output must be raised when individuals who sustained an uncomplicated mild TBI are in litigation or there is a potential for external incentive, because these contextual factors have been found to account for moderate effect sizes on negatively biased neuropsychological performance (Bianchini, Curtis, & Greve, 2006; Binder & Willis, 1991; Millis & Volinsky, 2001).

Malingering

Malingering has been defined as a negative response bias “designed to achieve some identifiable incentive” (Iverson & Binder, 2000, p. 832). The negative response bias can be in the form of exaggerated physical or psychological symptoms, while the external incentive can take the form of receiving financial compensation, avoiding military duty, avoiding work responsibility, obtaining drugs, escaping criminal prosecution, or evading liability in some way (APA, 2000; Slick, Sherman, & Iverson, 1999). With respect to TBI, malingering is typically manifested as diffuse and nonspecific neurological complaints, such as deficits in processing speed, memory, motor skills, sensation, abstract problem-solving skills, and fund of knowledge, as well as emotional disruption and non-epileptic seizures (Franzen & Iverson, 1998; Lynch, 2004). The key to a classification of malingering is that the exaggeration behaviour *must* be rational and intentional, and not due to an alternative explanation such as a psychiatric or neurological disorder (Iverson & Binder, 2000). Thus, when performing a differential diagnosis of TBI versus malingering, it is important to consider several other possible conditions that may be contributing to or exacerbating the observed behaviour. Explicitly, conversion disorder, factitious disorder, major depression disorder, and post-traumatic stress disorder (PTSD) tend to have profiles that overlap greatly with symptoms routinely observed in TBI as well as with deficits commonly exaggerated by malingerers. Factitious disorder, conversion disorder, and other somatoform disorders consist of symptom fabrication, but, unlike malingering, these disorders involve psychological rather than external incentives (APA, 2000; Iverson & Binder, 2000; Slick et al., 1999). In addition, conversion disorder and other somatoform disorders are distinguished from malingering and factitious disorder by the fact that the motivation of the behaviour is unconscious in nature (APA, 2000; Babin, 2002).

Major depression must be ruled out when performing differential diagnoses because it may mimic some aspects of malingering and TBI such as low motivation, lack of cooperation, apathy, negative views of personal functioning, long response lags, distractibility, cognitive slowing, and memory problems as well as headaches, excessive worry over health, and irritability (APA, 2000; Bordini, Chaknis, Ekman-Turner, & Perna, 2002; Iverson & Binder, 2000; Rao & Lyketsos, 2000). Lastly, individuals with PTSD tend to present with low cooperation, altered recall of aspects of the traumatic event, difficulties with concentration and memory, mood disturbance, irritability, absent “organic indicators,” avoidance behaviour, and other apparent inconsistencies that might be misinterpreted as displays of suboptimal effort or malingering (Bordini et al., 2002, p. 94). In the case of PTSD, a lack of purposeful exaggeration or fabrication of symptoms would differentiate it from malingering. Lastly, some of these disorders can actually present concomitantly with TBI (Rao & Lyketsos, 2000). Accordingly, an assessment of malingering should be very thorough because of the potentially aversive consequences carried by a misdiagnosis (e.g., delayed treatment) and because of the negative implications carried by such a label.

The DSM-IV-TR does not have a formal diagnosis for malingering; instead, malingering is in the "additional conditions that may be a focus of clinical attention" section (APA, 2000, p. 739). Although there were several tentative criteria for identifying malingering prior to 1999 (e.g., Greiffenstein, Baker, & Gola, 1994; Nies & Sweet, 1994; Pankratz, 1988), a uniform classification system was lacking, which prompted Slick et al. (1999) to formulate a set of comprehensive criteria based on test performance and specific contextual factors. Specifically, the proposed criteria for a diagnosis of Malingered Neurocognitive Dysfunction (MND) involve the presence of substantial external incentive (i.e., criterion A), as well as evidence of suboptimal performance from neuropsychological test data (i.e., criterion B) and/or self-report data suggestive of symptom exaggeration (i.e., criterion C) that cannot be due to psychiatric, neurological or developmental factors (i.e., criterion D).

Within criterion B, an individual can be classified as displaying a definite response bias (B1) if they perform below chance levels ($p < .05$) on tests specifically designed to detect feigned cognitive dysfunction, such as the Test of Memory Malingering (TOMM; Tombaugh, 1996). A classification of probable response bias (B2) results when the individual's performance is consistent with feigning on one or more well-validated psychometric tests or indices designed to measure exaggeration or fabrication of cognitive deficits (e.g., Digit Span-Vocabulary discrepancy; Mittenberg, Theroux-Fichera, Zielinski, & Heilbronner, 1995). An individual is also classified as exhibiting a probable response bias when there is a discrepancy between test data and known patterns of brain functioning (B3), observed behaviour (B4), the reports of trustworthy collaterals (B5), or documented background history (B6; Slick et al., 1999). Evidence satisfying criterion C can take the form of discrepancy between reported and documented history (C1); discrepancy between reported symptoms and known patterns of brain functioning (C2); discrepancy between reported symptoms and behavioural observations (C3); discrepancy between reported symptoms and information from collateral informants (C4); and evidence of exaggerated or fabricated psychological dysfunction (C5) from well-validated validity scales or indices from self-report measures, such as the Fake Bad Scale (Lees-Haley, English, & Glenn, 1991). Because there are no actual "gold standard" tests available to classify someone as malingering with 100% certainty, Slick et al. (1999) formulated their criteria to account for different degrees of certainty—namely, "possible MND", "probable MND", and "definite MND"—depending on the amount of evidence available. Consequently, these researchers have suggested using several sources of data as converging lines of evidence for a diagnosis of malingering (see Table 2 of Slick et al., 1999, for complete list of diagnostic criteria).

Unfortunately, but not surprisingly, it is very unlikely that an individual will divulge that they are feigning or exaggerating their symptoms during testing, especially when they are in litigation and there is a potential for secondary compensation (i.e., external incentive). Because of this absence of candour, actual prevalence rates of symptom over-reporting or feigning are unknown and must be estimated from base rates depending on the setting (Millis & Volinsky, 2001). Base rates are the number of cases that have been judged to be influenced by malingering from the total population given a specific setting. From different studies, the prevalence of suboptimal effort output has been reported as falling between 2-26% in clinical settings (Meyers & Volbrecht, 2003; Schretlen, 1988, as cited in Root, Robbins, Chang, & Van Gorp, 2006) and as ranging between 25% and 59% in forensic-based practices or in settings with potential for secondary gain (Greiffenstein & Baker, 2006; Greiffenstein et al., 1994; Larrabee, 2003a; Millis et al., 1995; Root et al., 2006). Slick, Tan, Strauss, and Hultsch (2004) conducted a recent survey on 24 neuropsychologists that specialize in detecting suboptimal effort output, which showed that the majority of them estimate the base rate of suboptimal effort output to be at least 10%. In comparison, Mittenberg, Patton, Canyock, and Condit (2002) surveyed 131 neuropsychologists and found that their estimates of suboptimal effort output ranged widely depending on their practice setting. Specifically, the base rates of probable malingering were noted as 8% for medical cases, 19% for criminal cases, 29% for personal injury claims, and 30% for disability claims. In the same vein, when Larrabee (2003a) pooled the results of 11 studies on suboptimal effort output, he found an average base rate of 40% for cases in neuropsychological settings with potential for secondary compensation. Thus, it appears that the prevalence of malingering may have been underestimated by the neuropsychologists surveyed by Slick et al. (2004).

Because of the high incidence of TBI in North America (CDC, 1999; Canadian Institute for Health Information, 2006) and because it is commonplace for individuals to seek some type of compensation following injury (Etcoff & Kampfer, 1996), there has always been a need to develop techniques to adequately assess the validity of reported symptoms. Thus, to address this issue, clinicians and researchers have employed several methods to assess insufficient effort and exaggerated responding. Apart from identifying the inconsistencies mentioned above, the most common performance-based approaches used to identify negative response bias consist of evaluating specialized tests of effort and atypical patterns of performance on conventional neuropsychological tests (Iverson & Binder, 2000; Millis & Volinsky, 2001; Slick et al., 1999; Slick et al., 2004). Prior to examining the different types of symptom validity measures and indices, however, it is important to address the terms commonly used to assess each test's validity. Essentially, a test or procedure is assessed by its diagnostic hit rate, sensitivity, specificity, and predictive power (Altman & Bland, 1994; Etcoff & Kampfer, 1996; Larrabee, 2003a; Mathias, Greve, Bianchini, Houston, & Crouch, 2002; Millis et al., 1995; Millis & Volinsky, 2001).

The *hit rate* is the total percentage of individuals correctly classified by the test. *Sensitivity* is the true positive rate for a test—i.e., the proportion of individuals supplying poor effort correctly classified as malingering. *Specificity*, on the other hand, is the true negative rate for a test—i.e., the percentage of individuals giving good effort correctly classified as not malingering. *Positive predictive power* (PPP) is calculated as the true positive value over the total number of individuals in the population receiving positive scores on a test (i.e., true and false positives; see equations A1 and A2 in the Appendix). It denotes the probability that an individual who received an abnormal test score on a test actually has the purported condition. In the case of malingering, PPP signifies the proportion of individuals receiving positive scores on a malingering test that were accurately labelled as malingerers. *Negative predictive power* (NPP) is calculated as the true negative value over the total number of individuals receiving negative scores (i.e., true and false negatives; see equations A3 and A4 in the Appendix). With respect to malingering, it represents the proportion of individuals receiving malingering-negative scores that were accurately labelled as non-malingerers. Both PPP and NPP are dependent on the accuracy of the test as well as the base rate of malingering in the population of interest (Heinly et al., 2005).

When developing tests or indices of insufficient effort, the goal is to achieve high sensitivity while minimizing the number of individuals falsely identified as malingering (i.e., reducing false positive errors). As mentioned, because there are no gold standards available with 100% sensitivity and 100% specificity, each clinician must decide the degree of accuracy with which to judge an individual’s effort output. Consequently, the general practice is to judge tests and their respective cut-offs as optimal if they produce false positive error rates less than or equal to 10% (Ashendorf, O’Bryant, & McCaffrey, 2003; Greiffenstein, et al., 1994; Mathias et al., 2002; Millis, 1992; Millis et al., 1995). Stated differently, the goal of effort-based diagnostic tests is to achieve at least 90% specificity. With respect to PPP, there are no specific cut-offs to determine what is adequate but there is a consensus that any value above 50% suffices because it indicates that there is more than .50 probability that the person is exaggerating their symptomatology (i.e., “more probable than not”; Curtis, Greve, Bianchini, & Brennan, 2006, p. 59; see also Heinly et al., 2005). Moreover, it is important to note that any test yielding a false positive error rate of 0% (i.e., 100% Specificity) is associated with a PPP of 100% regardless of the test’s sensitivity as long as its sensitivity is greater than zero (Coleman, Rapport, Millis, Ricker, & Farchione, 1998; Heinly et al., 2005; Millis, 2003).

Symptom validity testing. The vast majority of neuropsychologists (79%) recently reported using at least one symptom validity test (SVT) during their evaluations (Slick et al., 2004). SVTs are tools specifically designed to assess symptom exaggeration and negative response bias. Their format usually involves the presentation of some type of stimulus, which the participant must select after a delay from a series of forced choices with two alternatives (Lynch, 2004). Many of the items making up SVTs appear difficult at face value but are often very simple and largely unaffected by neurological disorders (Iverson & Binder, 2000; Millis et al., 1995). Consequently, malingerers tend to overestimate the degree to which they must feign their responses, resulting in poorer scores than those from patients with genuine head injuries or severe neuropathology, who actually tend to perform well (Frederick & Speed, 2007; Green, Allen, & Astner, 1996; Tombaugh, 1996).

There are two ways to analyze the results from SVTs in order to determine whether an individual is exaggerating symptoms or expending suboptimal effort—(1) evaluating whether the score is significantly below chance (Binder & Willis, 1991) or (2) evaluating whether the score is below an experimentally set cut-off derived from known samples (Iverson & Binder, 2000). Because performing significantly below chance is very uncommon and very unlikely to be due to variability in responding, such low scores tend to occur only when the examinee is purposely exhibiting a negative response bias, and receiving such scores are as close to a “gold standard” as malingering diagnostic tests get (Slick et al., 1999). However, using below chance cut-offs (i.e., 2 SDs) tends to yield high specificity at the expense of low sensitivity, resulting in many false negatives (Franzen & Iverson, 1998; Millis et al., 1995). Alternatively, the experimental cut-off approach consists of establishing floor performances for persons with brain injuries and then comparing an individual’s performance against said floor cut-off (Iverson & Binder, 2000). Scores below these floor cut-offs are suggestive of exaggerated symptom reporting because they are “inconsistent with the performance of that population” (p. 838). Although there are several SVTs available, the following three measures are well-validated and among the most widely used (Slick et al., 2004).

The Test of Memory Malingering (TOMM) is an SVT designed to assess memory complaints (Tombaugh, 1996). It consists of two consecutive trials in which the individual is presented with visual stimuli consisting of 50 line drawings of common objects followed by a two-choice discrimination task where they are required to pick the target response. If the individual identifies fewer than 45 items on trial 2, an optional retention trial is administered 15 minutes later without the benefit of another learning trial. Scores below 45 on trial 2 or on the retention trial are suggestive of suboptimal effort output. Through a series of studies, Tombaugh (1997) showed that using the criterion of 45 correct responses, the TOMM readily detected simulators told to exaggerate memory deficits (i.e., 100% sensitivity) while correctly classifying 99% of cognitively intact community dwellers and 95% of neurologically impaired outpatients, including patients with TBI, amnesia, and aphasia. In contrast, Tan et al. (2002) found lower classification rates (i.e., 74.1-80.8% sensitivity and 96.4% specificity) when using a simulation-based study design with the TOMM. Nonetheless, Tombaugh and others have shown that the TOMM is unaffected by differences in age, education, depression, psychosis, or severe neuropathology, and is only moderately affected by moderate to severe dementia (Duncan, 2005; Rees, Tombaugh, Gansler & Moczynski, 2002; Tombaugh, 1997).

Another commonly used two-alternative SVT is the Recognition Memory Test (RMT; Warrington, 1984). It consists of two 50-item subtests—one with faces as stimuli and the other with words—and it uses a recognition paradigm much like the TOMM. It has previously been used to detect potential malingerers complaining of mild TBI (Millis, 1992) as well as malingering simulators (Iverson & Franzen, 1994). The scores, however, may be affected by brain impairment, especially dementia (Strauss, Sherman, & Spreen, 2006).

The Word Memory Test (WMT; Green et al., 1996) is an SVT that was originally designed to assess verbal memory but that is frequently used to assess effort output because of its built-in indices of negative response bias (Iverson & Binder, 2000). The WMT consists of 20 semantically related word pairs that the examinee is required to recognize immediately after presentation while paired with foils (immediate recognition; IR), after a 30-minute delay while paired with new foils (delayed recognition; DR), and then from multiple choices (MC). Following these subtests, there is a paired-associates trial (PA), a delayed free recall procedure (DFR), and a long delay free recall procedure (LDFR), all of which are intended to evaluate verbal memory. Because memory-impaired individuals tend to complete the IR and DR subtests relatively easily, these—in conjunction with a consistency composite score (IR-DR consistency)—are used as measures of effort output (Green et al., 1996). Overall, the WMT has been shown to be very reliable in detecting malingering simulators (92.6% sensitivity and 100% specificity; Tan et al., 2002) as well as in differentiating between groups of TBI patients involved in litigation (Green, Iverson, & Allen, 1999). Recently, Flaro, Green, and Robertson (2007) found that the WMT was also sensitive in differentiating between low functioning parents trying to gain custody of their children and mild TBI patients with average intelligence seeking compensation. In fact, less than 2% of the former group performed below the effort cut-offs for IR, DR or IR-DR consistency, whereas 40% of the latter group failed the same cut-offs. Thus, the WMT shows merit as a measure of suboptimal effort output in litigants.

However, there are several caveats involved in using SVTs or any other measure specifically designed to assess suboptimal effort. First, they tend to be somewhat repetitious and lengthy which may cause some individuals to “become annoyed, stop attending, and, in so doing, perform poorly” (Bordini et al., 2002, p. 97). Second, although they are sensitive to obvious symptom magnification, they may not be sensitive enough to detect subtle or intricate malingering strategies or variations of effort output throughout the assessment (Franzen & Iverson, 1998). Moreover, because specialized tests of malingering tend to focus on one type of symptom exaggeration (e.g., memory), variable patterns of symptom exaggeration may go undetected (e.g., sensory loss, motor impairment, reduced processing speed, etc.; Meyers & Volbrecht, 2003). Another problem with some SVTs is that their cut-off scores are well-known in litigation, which makes them susceptible to coaching by attorneys (Ben-Porath, 1994; Gunstad & Suhr, 2001). Lastly, because of their face validity, SVTs can easily be identified as effort tests (Tan et al., 2002), especially if individuals are expecting the administration of such tests (Suhr & Gunstad, 2000), which may prompt them to employ more subtle and believable malingering strategies (Youngjohn, Lees-Haley, & Binder, 1999). As a result, over the past 10 to 15 years, there has been a push to investigate alternate and covert ways of assessing effort. The resulting method is to assess insufficient effort from patterns of performance from neuropsychological tests normally used in clinical practice; in particular, the drive has been to formulate and validate indices embedded in clinical tests.

Floor effect analyses. Apart from SVTs, another method of assessing suboptimal effort is known as the floor effect analysis. Along the lines of the second approach of interpreting SVTs, the principal feature in floor effect analyses is to compare the performance of a group of individuals with genuine head injuries to the performance of a group of individuals with questionable injuries and suboptimal effort (i.e., known-groups design; Millis & Volinsky, 2001). An important study exemplifying the floor effect method using a known-groups design was performed by Backhaus, Fichtenberg, and Hanks (2004), whereby the authors calculated the performance levels (i.e., 10th, 25th, and 50th percentiles) of a group of moderate-severe TBI outpatients on standard neuropsychological tests and designed cut-offs based on these scores. Then, they determined the classification accuracy of these floor cut-off scores by comparing the performance of a mild TBI group to that of a group of litigants with poor effort as classified by the RMT and the TOMM as well as by other Slick et al.'s (1999) criteria. In general, Backhaus et al. (2004) found that standard neuropsychological tests do a good job of distinguishing between patients with mild TBI and litigants putting forth insufficient effort when using at least a 50th percentile floor level cut-off. In fact, the PPP values for the tests used were all above 73.7% and NPP values ranged between 52.1% and 88.5% depending on the floor level and the base rate. More specifically, using a basal cut-off of 50th percentile, the results for some of the tests were: 48% sensitivity (84% specificity) for the Finger Tapping Test using the dominant hand (FTT-d) and 56% sensitivity (87.5% specificity) using the non-dominant hand (FTT-nd; Reitan & Wolfson, 1985); 40% sensitivity (92% specificity) for the Trail Making Test part A (TMT-A) and 56% sensitivity (80% specificity) for part B (TMT-B; Reitan & Wolfson, 1985); 56% sensitivity (100% specificity) for the written version of the Symbol Digit Modality Test (SDMT-W) and 64% sensitivity (96% specificity) for the oral version (SDMT-O; Smith, 1973 as cited in Backhaus et al., 2004); 80% sensitivity (84% specificity) for the Judgment of Line Orientation (JOLO; Benton, Hamsher, Varney, & Spreen, 1983); in addition to 68% sensitivity (80% specificity) for the Token Test (TT) and 76% sensitivity (84% specificity) for the Sentence Repetition (SR) subtests of the Multilingual Aphasia Examination (Benton & Hamsher, 1989) (see Table 4 of Backhaus et al., 2004, for complete data). Thus, it appears that using the 50th percentile cut-offs suggested by Backhaus et al. (2004) have good sensitivity when used to discriminate individuals putting forth suboptimal effort from those with mild TBI. However, given that the specificity values were calculated using a mild TBI group, applying the aforementioned cut-offs to a moderate-severe TBI group would likely result in lower specificity due to increased impairment. Thus, a prudent practice might be to use cut-off values at the 10th percentile when evaluating patients with moderate-severe TBI as a reference group in order to maintain specificity at 90%. Using such a cut-off on Backhaus et al.’s (2004) mild TBI sample would yield PPP values above 96% for all measures used. Thus, it appears that standard neuropsychological tests hold considerable promise in determining suboptimal effort output when well-validated cut-offs are used.

Embedded indices. Much like stand-alone SVTs, analyses using the floor effect approach have also been shown to be susceptible to coaching, as informed simulators tend to outperform naïve simulators on some standard neuropsychological measures

(Wise et al., 2006). An alternative method to the floor effect analysis is to examine an individual's performance on empirically-validated indices that are generated from parts of standard tests. This procedure has several advantages over using stand-alone SVTs. First, by using an embedded index, the clinician does not need to administer any additional tests, which subsequently makes the assessment battery more efficient by minimizing the total administration time (Meyers & Volbrecht, 2003). Second, because these indices are calculated from parts of standard tests, they are less obvious than SVTs, which makes them less susceptible to coaching (Coleman et al., 1998; Suhr & Gunstad, 2000). Moreover, having multiple check points during testing allows for validity assessment throughout the evaluation rather than solely at the beginning of the day, which is when most SVTs tend to be administered (Slick et al., 2004). Likewise, having multiple indices allows for assessment of negative response bias across several cognitive domains (e.g., processing speed, attention, etc.) rather than simply assessing exaggeration of memory deficits (Meyers & Volbrecht, 2003). Thus, because of the sheer number of potential indices that can be incorporated in a standard battery, it is more difficult for the dishonest responder to track their answers across tests, resulting in more inconsistencies and a higher likelihood that the clinician will suspect suboptimal effort output or exaggerated symptom reporting.

One of the earliest effort indices developed was derived from the Digit Span subtest of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981). In particular, Greiffenstein et al. (1994) devised the Reliable Digit Span (RDS), which consisted of the sum of the last forward string repeated with both trials correct and the last backward string repeated with both trials correct. Using a cut-off of 7, the RDS reliably distinguished between a group of individuals giving incomplete effort and groups with persistent post-concussive syndrome (68% sensitivity and 89% specificity) and severe TBI (70% sensitivity and 73% specificity). However, the “probable malingering” group used in this study was defined according to less stringent criteria than those proposed by Slick et al. (1999). As a follow-up using better defined groups, Mathias et al. (2002) found that the RDS adequately discriminated a group of individuals with external incentive giving suboptimal effort from a group of patients with various degrees of TBI without external incentive (i.e., 67% sensitivity and 93% specificity). In the same vein, Heinly et al. (2005) demonstrated that, using the standard cut-offs, the RDS detected 39% of individuals making up their MND group while correctly classifying 96% of their non-MND group. Moreover, these investigators found that the PPP values for RDS ranged between 52% and 91% when using base rates from 10 to 50% (Heinly et al., 2005).

Another index derived from the WAIS-R is the Vocabulary-Digit Span difference in scaled scores (V-DS; Mittenberg et al., 1995), which was formulated on the premise that large discrepancies between these subtests are rare. Using a discriminant function, they found that a V-DS discrepancy greater or equal to 2 detected 71% of malingering simulators while correctly classifying 79% of non-litigating patients with mild to severe head injuries. Millis, Ross, and Ricker (1998) confirmed the utility of V-DS to detect malingering (i.e., 79% sensitivity and 90% specificity) by comparing the performance of a sample of financially compensable mild TBI individuals giving incomplete effort against the performance of a group with moderate-severe TBI. Finally, Schwarz, Gfeller, and Oliveri (2006) demonstrated that, using the WAIS-III (Wechsler, 1997), the V-DS was sensitive enough to detect coached simulators (85.7% sensitivity) but at the expense of low specificity (i.e., 63.4%). However, this study was somewhat superficial as all groups were composed of randomly assigned undergraduate students, suggesting that the classification rates might have been different under more stringent and externally valid parameters.

One of the few effort indices derived from tests of cognitive flexibility and abstract reasoning is the Failure to Maintain Set (FMS) score from the Wisconsin Card Sort Test (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993). Using a cut-off score of greater than one, Suhr and Boyer (1999) found that the FMS correctly discriminated between “malingering” undergraduates and their “normal” counterparts (i.e., 70.7% sensitivity and 87.1% specificity) as well as between patients giving suboptimal effort and patients with mild to moderate TBI (i.e., 82.4% sensitivity and 93.3% specificity). Similar findings have been reported in other studies (Heinly, Greve, Love, & Bianchini, 2006; King, Sweet, Sherer, Curtiss, and Vanderploeg, 2002). In a study focussing solely on specificity using regression formulae including FMS and number of categories completed in the WCST, the values ranged between 98.7-100% in college students, 79.5% in substance abusers, 86.7-92.2% in mixed neurological samples, 88% in patients following stroke, and 58.3% in severe TBI patients (Greve & Bianchini, 2002).

Self-report measures of psychopathology—such as the Minnesota Multiphasic Personality Inventory, 2nd ed. (MMPI-II; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the Personality Assessment Inventory (PAI; Morey, 1991)—have also been studied as possible predictors of dissimulation. In particular, Lees-Haley et al. (1991) formulated the MMPI-II Fake Bad Scale (FBS) to detect malingering in personal injury litigants and found that it correctly classified 96% of them, while correctly identifying 90% of claimants with genuine injuries. This scale has proven valuable in differentiating between litigants exaggerating symptoms and persons with various degrees of TBI severity (i.e., 90% sensitivity and 90% specificity; Ross, Millis, Krukowski, Putnam, & Adams, 2004; see also Meyers, Millis, & Volkert, 2002) as well as between litigating and non-litigating patients with moderate-severe closed head injuries (i.e., 80.8% sensitivity and 86.2% specificity; Larrabee, 2003b; Larrabee, 2003c). In the same vein, the PAI Negative Impression Management scale (NIM) has been shown to distinguish between defendants malingering in court-ordered pre-trial evaluations (for competence and sanity) and honest responders (i.e., 91% sensitivity and 72% specificity; Boccaccini, Murrie, & Duncan, 2006).

Recently, Smith and Burger (1997) developed the Structured Inventory of Malingered Symptomatology (SIMS) to assess feigned or exaggerated psychiatric symptoms (e.g., depression) or cognitive dysfunction (e.g., memory loss). This specialized self-report SVT consists of five distinct scales encompassing commonly feigned conditions —Psychosis (P), Affective Disorders (AF), Low Intelligence (LI), Amnestic Disorders (AM), and Neurological Impairment (NI). Each scale consists of 15 non-overlapping items and has an independent cut-off to denote over-reporting of symptoms within that domain. Using a Dutch translation of the SIMS, Merckelbach and Smith (2001) demonstrated that it accurately distinguished undergraduate simulators from normal controls and psychiatric inpatients with PPP scores falling above 90%. On a separate analog study, Jelicic, Hessels, and Merckelbach (2005) used the SIMS to differentiate between honest responders (100% specificity), simulators asked to exaggerate psychotic symptomatology (93% sensitivity), simulators provided with information about psychotic symptoms (100% sensitivity), and simulators given information about psychotic symptoms and warned not to exaggerate (80% sensitivity). Although the overall hit rate was 94.6% and overall sensitivity was 91%, the authors used an unrealistic base rate of 75%, making the results difficult to generalize. Finally, Lewis, Simcox, and Berry (2002) used a known-groups design to assess the ability of the SIMS and MMPI-2 to differentiate between two groups of individuals participating in pre-trial psychological evaluations. These investigators found that both the SIMS and MMPI-2 validity scales yielded very high NPPs when the predetermined cut-offs were used (i.e., 100% and 92% respectively). Thus, although further research with this measure using known-groups designs is warranted, the SIMS appears effective for screening feigned reporting in forensic samples.

Lastly, Meyers and Volbrecht (2003) used patterns of performance from standard neuropsychological tests to distinguish between a group instructed to simulate neuropsychological impairment (i.e., analog design; Millis & Volinsky, 2001) and a group of individuals with varying degrees of TBI severity. In particular, Meyers and Volbrecht (2003) found that performance of any two measures below empirically derived cut-offs—including TT, SR, and JOLO—was suggestive of suboptimal effort output especially in the context of litigation. By means of similar methodology, Larrabee (2003a) used pair-wise combinations of test failures to discriminate between a group of patients with moderate-severe closed head injury and a group of definite MND individuals with potential for secondary gain. The indices and measures used included the RDS, FT, FMS, FBS, and visual form discrimination (VFD), which resulted in a classification accuracy of 91.6% (87.7% sensitivity and 94.4% specificity). In a follow-up study using likelihood ratios, Larrabee (2008) demonstrated that failing two or three SVTs resulted in very good PPPs (ranging from 73.5% to 99.9%) depending on the base rate used (i.e., 10 to 90%). In general, consistent with Meyers and Volbrecht (2003), Larrabee (2003a, 2008) found that using multiple indicators to determine suboptimal effort resulted in much higher classification rates than using a single measure. Thus, defining suboptimal effort as two or more failed effort measures appears to distinguish effectively between individuals feigning or exaggerating symptoms and individuals with genuine moderate-severe TBI.

Taken as a whole, indices embedded in standard neuropsychological measures appear to be worthwhile in detecting effort output, especially in litigants. These measures are not as sensitive as stand-alone effort tests but because the former do not require additional administration time, they are more efficient and allow the clinician to make preliminary diagnostic decisions quickly in order to maximize the utility of the assessment. In addition, these embedded indices have low face validity with respect to effort detection because they are derived from actual clinical tests measuring cognitive performance, which subsequently makes them less susceptible to coaching practices. Lastly, using embedded indices throughout the examination and across modalities makes it more difficult for a dishonest responder to keep track of their exaggerated response style, thereby increasing the likelihood that their negative response bias will be detected. Of all tests used to derive embedded effort indices, the CVLT and CVLT-II (discussed in the next section) appear to hold the most promise because of their adequate difficulty level, multiple built-in validity scales, and serial testing, which affords the clinician multiple variables and testing points to assess not only different aspects of learning and memory but also effort output.

CVLT and CVLT-2 research. The CVLT (Delis, Kramer, Kaplan, & Ober, 1987) and its successor, the CVLT-II (Delis et al., 2000), are popular clinical tests of verbal learning, encoding, and retrieval. Although both editions share the majority of test components (i.e., 5 learning trials, a distractor list, immediate and delayed free and cued recall trials, and a delayed yes/no recognition trial), the CVLT-II has new word lists that are easier to comprehend, a better conceptualized yes/no recognition trial, additional validity measures, and a larger normative sample (Delis et al., 2000; Strauss et al., 2006). A full discussion of the CVLT components and procedures falls outside the scope of this paper; consequently, only the CVLT-II will be described in detail. Nonetheless, a review of research on malingering involving both the CVLT and the CVLT-II will follow.

The CVLT-II consists of a list of 16 words—comprised of 4 words for each of 4 semantic categories (i.e., animals, furniture, modes of transportation, and vegetables)—that are initially read 5 times while the individual recalls as many words as they can immediately following each presentation (i.e., list A ; trials 1-5). Then, the individual is provided with a distractor list of 16 different words (i.e., list B), which they are supposed to recall immediately after their presentation. After this task, they must recall as many words as they can from list A without and with semantic cueing (i.e., Short-Delay Free Recall and Short-Delay Cued Recall, respectively). After a 20-minute delay filled with tasks assessing other domains (i.e., not memory or verbal material), the participant is asked to recall as many words as they can from list A without and with semantic cueing (i.e., Long-Delay Free Recall and Long-Delay Cued Recall, respectively). Following the delayed cued recall trial, they are presented with a yes/no recognition task composed of the 16 target items from list A as well as 32 additional foils made up of four different categories—UN, BN, PR, and BS. The UN variable consists of novel false positive foils that are semantically unrelated to the previously presented list A items. The BN group of foils consists of words that were presented as part of the distractor list (i.e., list B) but that are not semantically related to any of the list A items. The PR foils are novel but semantically or "prototypically" related to the target items. Lastly, the BS group consists of foils that were presented as part of the distractor list and that are semantically related to target items. Finally, an optional forced-choice recognition task may be administered to the individual after a 10-minute delay in order to screen for effort output (Delis et al., 2000; Donders & Moore, 2004; Root et al., 2006).

Although not typically generated from the CVLT-II, the yes/no recognition foils lend themselves to be pooled into two separate composite scores based on their categorical properties. The first composite is made up of items semantically unrelated to the target items (i.e., pooled from the UN and BN categories) and can be considered "easy to reject" (ETR) because its foils are not part of any of the target categories (i.e., animals, furniture, modes of transportation, and vegetables), which makes them more easily discernable from target items. Contrastingly, the second composite consists of items that are semantically related to the target items (i.e., pooled BS and PR items) and are consequently hypothesized to pose more difficulty to inhibit for the participant (i.e., "difficult to reject"; DTR).

Overall, the output generated from a CVLT or CVLT-II protocol yields over 30 scores and indices—many of which have been studied as potential indicators of suboptimal effort (Millis et al., 2007). Several researchers (Ashendorf et al., 2003; Coleman et al., 1998; Curtis et al., 2006; Demakis, 1999; Millis et al., 1995; Millis et al., 2007; Moore & Donders, 2004; Root et al., 2006; Sweet et al., 2000) have conducted many different studies examining the utility of variables from the CVLT and CVLT-II in detecting insufficient effort. The CVLT and CVLT-II variables that have been studied most extensively include: total number of words learned across five trials (Total 1-5), Short-Delay Free Recall (SDFR), Short-Delay Cued Recall (SDCR), Long-Delay Free Recall (LDFR), Long-Delay Cued Recall (LDCR), Recognition Hits (Hits), Recognition Discriminability (*d'*), Forced-Choice Recognition total (FCR; CVLT-II only), and Critical Item Analysis (CIA; CVLT-II only). Of these variables, those derived from the yes/no recognition trial have shown the most consistency in detecting suboptimal effort. More specifically, Millis et al. (1995) performed both linear and quadratic discriminant functions using Total 1-5, Hits, *d'*, and LDCR from the CVLT and found that either discriminant function yielded good accuracy in differentiating a moderate-severe TBI group from a mild TBI group with insufficient effort (i.e., 83-96% sensitivity and 91-96% specificity). Then, they conducted univariate frequency distributions to determine adequate cut-offs for each variable (set at a maximum of 10% false positive rate) and found that *d'* and Hits yielded very good accuracy rates that were comparable to the more complex discriminant function, whereas the Total 1-5 and LDCR variables yielded lower, but still suitable, rates. These authors also examined the pattern of hits and false-positive errors in the yes/no recognition trial to determine whether the *d'* variable was sensitive to more than one type of malingering strategy—denial response style (low hits, low false positives), combination (low hits, high false positives), and “yes” response bias (high hits, high false positives). Of the participants in the group putting forth insufficient effort, 12, 10, and 1 participant used the above strategies, respectively. Thus, the Millis et al. (1995) study gave some preliminary evidence for the presence of different malingering strategies used in the yes/no recognition trial. Sweet et al. (2000) expanded on Millis et al.'s (1995) study by adding malingering simulators and normal controls to the TBI group and the group of clinical malingerers. With respect to the classification accuracy that resulted from the study by Sweet et al. (2000), the specificity values were comparable to the Millis study (i.e., Total 1-5: 76-100%; LDCR: 74-95%; Hits: 83-100%; *d'*: 81-100%) but the sensitivity values were slightly lower (i.e., Total 1-5: 52-80%; LDCR: 48-62%; Hits: 48-88%; *d'*: 57-68%), depending on the group and cut-offs used. The classification rates found in the Sweet study were comparable to those found by Ashendorf et al. (2003) as well as Curtis et al. (2006) when these authors assessed the classification accuracies of the aforementioned variables as well as several regression models and the Millis discriminant function.

With respect to the CVLT-II, Bauer, Yantz, Ryan, Warden, and McCaffrey (2005) constructed a discriminant function using Total 1-5, *d'* scaled score, Hits, LDCR, and FCR in an attempt to differentiate between two groups of patients with mild-moderate TBI—one putting forth adequate effort and one giving insufficient effort. Although the authors found specificity scores similar to those found using the CVLT (i.e., 95.6%), sensitivity was much lower than in previous research (i.e., 13.8%). To address the need for new CVLT-II validity indicators, Millis et al. (2007) used Bayesian model averaging (BMA), a multivariate logistic regression model, to investigate which of 18 CVLT-II variables best distinguished between persons with moderate-severe TBI and litigants with mild TBI supplying insufficient effort. From the many models resulting from the BMA, posterior probabilities above .50 were only found for LDFR (.597), *d'* (.924), and Recall Discriminability standard score (1.00), suggesting that these were the best predictors of suboptimal effort. Based on these variables, the authors performed a logistic regression analysis to determine their utility as validity indicators and to formulate a function on which individuals' scores could be entered to determine the probability of suboptimal effort.

Finally, Root et al. (2006) used the FCR and CIA to differentiate between individuals referred for clinical evaluations, individuals referred for forensic evaluations showing adequate effort, and individuals referred for forensic evaluations putting forth suboptimal effort. Briefly, the FCR and CIA are concise screens of effort output built into the CVLT-II. In the FCR, the examiner reads a list of 16 word pairs (original word accompanied by a novel and semantically unrelated word) and participant must select the word that belonged to the original list. This task is very simple—90% of the participants in the normative sample received a perfect score (Delis et al., 2000). The CIA is a consistency measure comprised of two indices—one that examines whether an item was previously recalled at least once during the recall trials (CIA-recall) but not during the FCR trial, and another that evaluates whether an item was recalled during the yes/no recognition trial but not during the FCR trial (CIA-recognition). In their study, Root et al. (2006) found that using a cut-off value below 16 correct on the FCR (i.e., less than 100%) resulted in 60% sensitivity and 81% specificity, whereas using a cut-off value of at least one yielded 36% sensitivity and 78% specificity on CIA-recall, and 32% sensitivity and 81% specificity on CIA-recognition. Overall, incomplete effort appears to be associated with poor performance on recognition tasks on the CVLT-II, although sensitivity values vary considerably between studies.

Within the yes/no recognition trial of the CVLT-II, the total number of false positive errors has not been studied exclusively with respect to suboptimal effort output. In fact, Millis et al.'s (2007) BMA showed that this variable was seldom present in their prediction models. However, because *d'*—a statistic combining both hit rate and false positive rate—consistently emerges as a variable that discriminates between moderate-severe TBI and suboptimal effort, some component of *d'* (i.e., false positive errors or misses) must account for its predictive power. More specifically, given that *d'* has previously yielded better classification accuracy than Hits (i.e., 16 - misses) using both univariate (Ashendorf et al., 2003; Millis et al., 1995) and multivariate approaches (Coleman et al., 1998), suggests that false positive errors (i.e., endorsed foils) may also play a role in the predictive power of *d'*.

As mentioned above, the recognition foils can be divided into four distinct categories based on their novelty of presentation and semantic relation to the target items (i.e., UN, PR, BN, and BS) as well as two composites based solely on semantic relation to target items (i.e., ETR and DTR). Because the different types of foils and composites appear discrepant in difficulty level, exploring patterns of responses may prove useful in detecting suboptimal effort output. In particular, those foils semantically unrelated to the target items (i.e., ETR) should be easily discernable from target items, whereas those with semantic relation to the targets (i.e., DTR) should be more difficult to inhibit. Within the semantically unrelated items (i.e., ETR composite), those foils that are novel (i.e., UN foils) should be easier to discriminate from the target items than those previously presented as part of list B (i.e., BN foils) because memory traces of the latter items may increase the chances of their endorsement during the yes/no recognition trial. To this end, this study examines the patterns of endorsement of recognition foils in litigants reporting mild TBI and showing non-credible performance versus the pattern of endorsement in patients with moderate-severe TBI. Within a paradigm of discrepant item difficulty, it is expected that those attempting to exaggerate symptoms would not only indiscriminately endorse a higher number of recognition foils than patients with genuine head injuries but also endorse items that are seldom endorsed by these clinical populations. Such a response style would not be in keeping with expected levels of brain functioning since patients with moderate-severe TBI have been shown to have similar retrieval abilities to demographically-matched healthy controls on the CVLT yes/no recognition trial (Vanderploeg, Crowell, & Curtiss, 1999). Moreover, Baldo, Delis, Kramer, and Shimamura (2002) recently found that although patients with focal frontal lobe lesions made considerably more errors than healthy controls in the yes/no recognition subtest of the CVLT-II, the groups did not differ with respect to the rate of endorsement of UN items. Thus, as a natural extension of the extant literature, participants in the moderate-severe TBI group in the present study are expected to endorse fewer semantically unrelated foils than those in the MND group. Moreover, it is expected that responders purposefully attempting to misrepresent themselves as having severe neuropathology will endorse novel, semantically unrelated items. Stated differently, it is hypothesized that the MND group will endorse (1) more UN foils than any other foil; and (2) more ETR items than DTR items compared to the sample with moderate-severe TBI.

Method

Participants

Two groups of demographically-matched participants were studied retrospectively using data retrieved from 282 consecutive admissions to the Rehabilitation Psychology and Neuropsychology department at the Rehabilitation Institute of Michigan (RIM: Detroit, MI) and 41 medico-legal cases from Psychological Systems Inc. (PSI: Royal Oak, MI) from 2000 to 2007. The malingered neurocognitive dysfunction (MND)[1] group consisted of 31 outpatients (20 male, 11 female) between the ages of 22 and 71 (*M* = 42.29; *SD* = 12.39) and ranging in education from 8 to 18 years (*M* = 12.65; *SD* = 2.65; see Table 1 for descriptive statistics).

Table 1

Descriptive Statistics of the Participants Making up the Reference and Criterion Groups

	MND (*n* =31)		TBI (*n* = 82)		T-Test		
Variable	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
Age	42.29	12.39	41.44	12.37	-0.33	111	.74
Years of education	12.65	2.65	12.39	2.09	-0.54	111	.59
Months post injury	56.39	49.43	75.67	58.68	1.76	63.78[a]	.08

Note. [a]Equal variances not assumed. MND = malingered neurocognitive dysfunction group; TBI = moderate-severe traumatic brain injury group.

All individuals in the MND group were evaluated at an outpatient clinic after reporting significant impairments associated with a head injury. Twenty five of the 31 participants were medico-legal cases from PSI whereas 6 participants came from the RIM outpatient database. Of the participants making up the MND group, 24 were in motor

---

[1] The term “malingered neurocognitive dysfunction (MND)” is used to maintain continuity with the term used by Slick et al. (1999). In the context of this study, MND is operationally defined as a negative response bias on neuropsychological testing or symptom exaggeration on self-report measures in the presence of an identifiable and substantial secondary gain (i.e., personal injury or worker’s compensation). The performance is not in keeping with expected levels of functioning and not attributable to psychiatric, neurological, or developmental factors.

vehicle accidents (MVA), 4 were pedestrians in MVAs, 1 was riding a motorcycle in an MVA, 1 was involved in a fall, and 1 suffered an anoxic event. All MND participants claimed they were unable to maintain employment due to the cognitive impairment resulting from their accident or event. All individuals making up the MND group were actively pursuing personal injury or worker's compensation at the time of assessment.

Nineteen participants had very brief (i.e., < 5 minutes) or no reported LOC, while 3 individuals experienced LOC lasting between 5 and 30 minutes. No data on LOC were available for 9 participants. PTA had been monitored prospectively. Twelve participants never experienced PTA, while 4 participants displayed PTA lasting less than 1 hour, and 15 participants had unknown PTA status. The groups were also classified using the lowest GCS score on file. Twenty participants scored above 13 on the GCS, whereas GCS data was unknown or the test not administered for the other 11 participants. All MND participants had normal CT and/or MRI scans, and displayed no focal neurological deficits on neurological examination. When data were unknown with respect to any of the aforementioned acute injury staging variables, a participant was required to have met the other three criteria in order to be included in the MND group. All MND participants were negative with respect to major psychiatric history, history of psychiatric hospitalization, comorbid psychological or psychiatric disorders, and learning or developmental disabilities. Thus, overall, the selection criteria used in this study were comparable to, if not more stringent than, those used by Dikmen et al. (2001) to define mild TBI (i.e., LOC < 1 hour, PTA < 24 hours, GCS > 12, and negative neuroimaging). Given these inclusion criteria, all participants were considered to have sustained *at most* mild TBIs. Moreover, because all MND participants were evaluated an average of 56.39

months (*SD* = 49.43) post-injury, no residual prolonged post-injury impairment was expected in the majority of participants.

All 31 MND participants were determined to be putting forth suboptimal effort during the assessment based on Slick et al.’s (1999) criteria. More specifically, each MND participant performed below well-validated cut-offs on a minimum of two tests designed to detect negative response bias or symptom exaggeration (i.e., criterion B2) or indices derived from self-report measures suggesting exaggerated psychopathology (i.e., criterion C5; Slick et al., 1999). One participant also performed significantly below chance levels on the TOMM to satisfy criterion B1 (see Table 2 for a list of measures and respective cut-offs used to determine effort output and criterion met). Criteria B1, B2, and C5 were selected to determine MND status because they were easily quantified from test performance and, thus, appeared more objective than criteria based on observations or collateral information. Of the 31 participants in the MND group, 23 “failed” at least three SVTs or indices, and 14 “failed” at least four SVTs or indices (*M* = 3.71; *SD* = 1.70; see Table 3 for a breakdown of the number of effort measures failed by participants in each group). Consequently, based on these selection criteria, all participants in this group were classified as *Probable* or *Definite MND* as determined by Slick et al.’s (1999) guidelines.

Table 2

Measures and Indices Used to Determine Effort Output in Cases and Slick et al.'s (1999) Criteria Met by Cases in Each Group

Measure or index	Cut-off score	Source	Criterion	Number of cases meeting criterion (%)[a]	
				MND (*n*=31)	TBI (*n*=82)
TOMM	Trial 2 or retention <19	Tombaugh (1996)	B1	1/27(3.7)	0/57(0)
WMT	IR, DR, or IR-DR <37.5%	Green et al. (1996)		0/13(0)	0/16(0)
TOMM	Trial 2 or retention <45	Tombaugh (1996)	B2	18/27(66.7)	0/57(0)
WMT	IR, DR, or IR-DR <82.5%	Green et al. (1996)		12/13(92.3)	0/16(0)
RMT-W	<=33	Iverson & Franzen (1994)		1/4(25)	-
RMT-F	<=30	Iverson & Franzen (1994)		2/4(50)	-
RDS	<7	Greiffenstein et al. (1994)		11/18(61.1)	0/82(0)
V-DS	>=|2|	Mittenberg et al. (1995)		7/10(70)	0/82(0)
VFD	<26	Larrabee (2003a)		3/4(75)	0/8(0)
JOLO	<=12	Meyers & Volbrecht (2003)		0/3(0)	0/23(0)
SDMT-W	<30	Backhaus et al. (2004)		12/19(63.2)	4/79(5.1)
SDMT-O	<35	Backhaus et al. (2004)		13/19(68.4)	0/79(0)
TMT-A	>63.3	Backhaus et al. (2004)		9/30(30)	2/81(2.5)
TMT-B	>192.6	Backhaus et al. (2004)		8/30(26.7)	3/79(3.8)
TT	<40	Backhaus et al. (2004)		2/2(100)	0/17(0)
SR	Adjusted score <9	Backhaus et al. (2004)		1/1(100)	-
FTT-d	<37.7	Backhaus et al. (2004)		4/7(57.1)	0/39(0)
FTT-nd	<23.5	Backhaus et al. (2004)		0/7(0)	0/39(0)
FMS	>1	Suhr & Boyer (1999)		2/2(100)	11/11 (100)
FBS	>21	Ross et al. (2004)	C5	6/8(75)	0/27(0)
NIM	*T*>80	Boccaccini et al. (2006)		2/3(66.7)	0/9(0)
SIMS	NI>2 or AM>2 or Total>14	Widows & Smith (2004)		2/2(100)	-

Note. All cut-off scores are raw values unless stated otherwise. [a]Participants were not administered every measure. AM = amnestic disorders scale; DR = delayed recognition; FBS = fake bad scale; FMS = failure to maintain set; FTT-d = Finger Tapping Test - dominant hand; FTT-nd = Finger Tapping Test - nondominant hand; IR = immediate recognition; IR-DR = IR-DR consistency; JOLO = Judgment of Line Orientation; MND = malingered neurocognitive dysfunction group; NI = neurological impairment scale; NIM = negative impression management index; RDS = reliable digit span; RMT-F = Recognition Memory Test - faces; RMT-W = Recognition Memory Test - words; SDMT-O = Symbol Digit Modalities Test - oral; SDMT-W = Symbol Digit Modalities Test - written; SIMS = Structured Inventory of Malingered Symptomatology; SR = sentence repetition; TBI = moderate-severe traumatic brain injury group; TMT-A = Trail Making Test - part A; TMT-B = Trail Making Test - part B; TOMM = Test of Memory Malingering; TT = token test; V-DS = vocabulary-digit span discrepancy; VFD = Visual Form Discrimination; WMT = Word Memory Test.

Table 3

Number of Effort Measures Failed by Participants in Each Group

Effort measures failed	MND (*n*=31)	TBI (*n*=82)
0	0	62
1	0	20
2	8	0
3	9	0
4	5	0
5	7	0
6	0	0
7	0	0
8	1	0
9	1	0

Note. MND = malingered neurocognitive dysfunction group; TBI = moderate-severe traumatic brain injury group.

The data for the participants making up the moderate-severe TBI group were retrieved from the SEMTBIS database at RIM. To be accepted into the SEMTBIS program, individuals met the diagnosis of TBI (including penetrating wounds) as evidenced by LOC due to brain trauma, or PTA, or skull fracture, or objective neurological findings that could be reasonably attributed to TBI on the initial physical examination or mental status examination within 72 hours of injury (SEMTBIS, 2004). The moderate-severe TBI group was composed of 82 outpatients (65 male, 17 female) between the ages of 17 and 69 (*M* = 41.44; *SD* =12.37) and ranging in education from 8 to 18 years (*M* = 12.39; *SD* = 2.09). An independent samples *t*-test revealed no significant differences between the TBI and MND groups for age, education, or months post-injury (see Table 1). In addition, chi-square analyses revealed that the two groups did not differ in regards to gender proportions, χ^2 (1, *N* = 113) = 1.90, *p* = .17 (Yates' Correction for Continuity), or in proportion of different ethnicities, χ^2 (5, *N* = 113) = 8.05, *p* = .15.

A mechanism of injury was present in all participants with moderate-severe TBI. Of the patients making up the moderate-severe TBI group, 30 sustained trauma from blunt assaults, 25 were in MVAs, 9 were involved in falls, 8 suffered gunshot wounds, 6 were pedestrians in MVAs, and 4 were on motorcycles when the MVAs took place. Of all these patients, 20 displayed moderate LOC (1-24 hours), 43 had severe LOC (1-7 days), 14 showed very severe LOC (1-4 weeks), 3 displayed extremely severe LOC (> 4 weeks) and 2 had unknown LOC periods, as determined by Jennett's (1979) brain injury classification guidelines. Three participants experienced PTA lasting between 1 and 24 hours (Moderate), 12 displayed PTA lasting up to a week (Severe), 35 had PTA lasting between 1-4 weeks (Very Severe), 21 had PTA lasting longer than 4 weeks (Extremely Severe) and 11 had a PTA period that was present, but undetermined. Thirty eight patients received GCS scores between 9 and 12 (i.e., moderate), 42 scored between 3 and 8 (i.e., severe), and 2 were in a chemically-induced coma and/or intubated within the first 72 hours post-injury so their GCS scores were unknown. Of the patients making up the moderate-severe TBI group, 70 had visible pathology on neuroimaging or showed positive neurological signs. The breadth of pathology, pathological signs, and associated sequelae resulting from the TBI included pupillary abnormalities (*n* = 24), cranial nerve involvement (*n* = 14), intracranial compression (*n* = 25), intracranial hypertension (*n* = 13), herniation syndrome (*n* = 2), seizures (*n* = 11), skull fractures (*n* = 12), CSF leaking (*n* = 2), craniotomy (*n* = 1), hydrocephalus (*n* = 1), and intraventricular, petechial, or subarachnoid haemorrhages (*n* = 39). Consequently, all TBI participants were deemed to have sustained *at least* moderate to severe TBIs. All TBI patients volunteered to be tested for research purposes and were gauged as putting forth adequate effort. Moreover, only cases failing fewer than two effort measures (*M* = 0.26; *SD* = 0.44) were included in the moderate-severe TBI group (see Table 3). No case was in active litigation or pursuing worker's compensation at the time of testing (see Table 4 for a summary of the inclusion criteria).

Table 4

Inclusion Criteria for Groups

Criterion	MND (*n* =31)	TBI (*n* = 82)
Quantifiable brain trauma	no	yes
External incentive (i.e., litigation, worker's comp.)	yes	no
Failed at least two effort measures	yes	no
Ruled out psychiatric or developmental factors	yes	yes

Note. MND = malingered neurocognitive dysfunction group; TBI = moderate-severe traumatic brain injury group.

The rationale behind generating such markedly discrepant MND and moderate-severe TBI criterion groups was to provide unequivocal external criteria for the formulation of the symptom exaggeration scales, as described below. In other words, any differences in test performance whereby the moderate-severe TBI group outperforms the MND group cannot be due to the latter group's severity of impairment. Instead, such discrepancies would likely be due to external sources, such as suboptimal effort.

Materials and Procedure

The CVLT-II was administered in a standard fashion (Delis et al., 2000), as part of a comprehensive neuropsychological battery. Some participants did not complete every test in the battery due to the severity of their impairment, fatigue, time constraints, and other such testing obstacles. Overall, all participants with moderate-severe TBI were administered at least five of the effort indices and measures and the participants making up the MND group were administered at least eight of the effort measures and indices. All standard tests and procedures were administered by trained psychometrists.

The CVLT-II variables of interest in this study were the number of items endorsed from each of the different types of yes/no recognition foils (i.e., UN, BN, PR, and BS) as well as the total number of items endorsed from the two composite variables (i.e., ETR and DTR). In addition to the original sample, these predictors were used to analyse a sub-sample of closed-head injury participants from the larger original study sample in order to determine whether there were any differences in the classification results when open head injuries are ruled out. Besides using the aforementioned yes/no recognition foils and composites, this study also sought to replicate previous malingering research involving the CVLT and CVLT-II yes/no recognition trial via independent multivariate analyses of Total 1-5, Hits, LDCR, and *d'* (Millis et al., 1995; Sweet et al., 2000) and LDFR, *d'*, and Recall Discriminability Scales Score (Millis et al., 2007), as well as a univariate analysis using Hits as a predictor (Coleman et al., 1998; Curtis et al., 2006). Finally, the CVLT-II variables used in this study (UN, BN, PR, BS, ETR, and DTR) and their respective diagnostic cut-offs were validated with two additional samples—patients with complicated mild TBI giving good effort (MTBI) and patients with complicated mild TBI putting forth suboptimal effort (SE)—in order to assess the generalizability of the findings.

Results

Means and standard deviations for the selected CVLT-II variables appear in Table 5 as raw scores. As predicted, the moderate-severe TBI group endorsed significantly fewer items from the UN variable and ETR composite than the MND group. In addition, the MND group endorsed more BN foils and generally made more false positive errors than the moderate-severe TBI group, whereas the moderate-severe TBI group endorsed more recognition hits than the MND group. Consequently, recognition discriminability was significantly lower for the MND group than the moderate-severe TBI group (see Table 5).

Table 5

Mean Raw Scores and Test Statistics for Selected CVLT-II Variables

	MND (n = 31)		TBI (n = 82)				
CVLT-II Variable	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
UN	1.52	1.48	0.10	0.30	-5.30	30.93[a]	.00**
BN	1.87	1.93	0.82	1.32	-2.81	41.03[a]	.00*
PR	2.23	2.04	2.02	1.85	-0.50	111	.62
BS	2.03	1.66	1.54	1.75	-1.36	111	.18
ETR	3.39	3.07	.91	1.42	-4.31	34.92[a]	.00**
DTR	4.26	3.43	3.56	3.19	-1.02	111	.31
False Positive Errors	7.65	6.18	4.48	4.12	2.64	40.47[a]	.01
Hits	10.74	2.65	13.61	2.27	-5.72	111	.00**
d’	1.44	1.08	2.43	.87	-5.01	111	.00**

Note. All values are raw scores unless stated otherwise. [a]Equal variances not assumed. BN = semantically unrelated foils from list B; BS = semantically related foils from list B; CVLT-II = California Verbal Learning Test – Second Edition; *d’* = recognition discriminability; DTR = difficult to reject foils; ETR = easy to reject foils; MND = malingered neurocognitive dysfunction group; PR = novel and prototypically related foils; TBI = moderate-severe traumatic brain injury group; UN = novel and semantically unrelated foils.

*$p < .01$ **$p < .001$.

Logistic Regression Analyses for CVLT-II Foils

The relationships between each pair of CVLT-II predictors and the criterion variable were investigated using Pearson product-moment correlation coefficients. Preliminary analyses were performed to ensure no violation of normality, linearity, and homoscedasticity. Of all the predictors, only UN, BN, and ETR were significantly correlated to the criterion variable (see Table 6). There were moderate to strong correlations found among all predictors. Very strong positive correlations were observed between the composite scores and their respective components (i.e., $r > .80$). Given such high correlations, multiple logistic regression analyses were conducted separately for the set of predictors and the composite scores in order to avoid multicollinearity (Tabachnick & Fidell, 2001; S. R. Millis, personal communication, July 20, 2007).

Table 6

Bivariate Correlations of CVLT-II Predictors and Criterion Variable

	UN	BN	PR	BS	ETR	DTR	Criterion
UN	1	.52**	.47**	.43**	.81**	.50**	.62**
BN	.52**	1	.46**	.48**	.92**	.52**	.30*
PR	.47**	.46**	1	.60**	.53**	.90**	.05
BS	.43**	.48**	.60**	1	.52**	.88**	.13
ETR	.81**	.92**	.53**	.52**	1	.59**	.49**
DTR	.50**	.52**	.90**	.88**	.59**	1	.10
Criterion	.62**	.30*	.05	.13	.49**	.10	1

Note. BN = semantically unrelated foils from list B; BS = semantically related foils from list B; CVLT-II = California Verbal Learning Test – Second Edition; Criterion = group membership (i.e., TBI or MND); DTR = difficult to reject foils; ETR = easy to reject foils; PR = novel and prototypically related foils; UN = novel and semantically unrelated foils.

*$p < .01$, 2-tailed. **$p < .001$, 2-tailed.

A direct multiple logistic regression analysis was performed on MND status as outcome and four predictors based on the CVLT-II yes/no recognition foils: novel-semantically unrelated items (UN), novel-semantically related (PR), list B-semantically unrelated (BN), and list B-semantically related (BS). Preliminary analyses revealed adequate sample size-to-predictors ratio (Millis, 2003).

A test of the full model with all the predictors against the constant-only model yielded statistically significant results, χ^2 (4, N = 113) = 73.18, $p < .001$, indicating that the predictors, as a set, reliably distinguished between patients with quantifiable moderate-severe TBI and litigants putting forth suboptimal effort. Based on the Hosmer and Lemeshow Goodness of Fit Test, the model had good fit ($p = .85$), indicating that model prediction did not differ significantly from observed values. Inspection of residuals revealed three outliers—all of which were participants from the MND group that the model misclassified as belonging to the moderate-severe TBI group. These outliers were not removed because the model had good fit even with them included in the analysis. The model showed adequate convergence and the standard errors for the parameters were not exceedingly large; thus, there was no multicollinearity evident (Tabachnick & Fidell, 2001). The full model evidenced good improvement over the constant-only model ($-2LL_0 = 132.78$; $-2LL_{FULL1} = 59.60$; Cox & Snell $R^2 = .48$; Nagelkerke $R^2 = .69$). Using a .50 classification cut-off, the model correctly classified 88.5% of the participants, with 61.3% sensitivity and 98.8% specificity. The PPP was 95.0% and the NPP was 87.1%, suggesting very high predictive power (see Table 7). Leave-one-out cross-validation resulted in correct classification of 86.7% of the grouped cases, indicating excellent generalizability.

Table 7

Classification Table for Full Model Using All the Foils

		Predicted	
		TBI	MND
Observed	TBI	81	1
	MND	12	19

Note. MND = malingered neurocognitive dysfunction group; TBI = moderate-severe traumatic brain injury group.

Table 8 shows regression coefficients, Wald statistics, odds ratios, and 95% confidence intervals for odds ratios for each of the four predictors. According to the Wald criterion, the UN foils and the PR foils reliably predicted a person’s effort output during testing. However, given that PR emerged as a significant predictor in the multivariate regression model while having a near-zero (and not significant) correlation with the criterion variable suggests that it may have acted as a suppressor variable in the regression equation. Consistent with this contention is the fact that PR’s beta weight and its correlation with the criterion variable have different signs (Tabachnick & Fidell, 2001).

Table 8

Logistic Regression Analysis of MND Status as a Function of CVLT-II Foils

CVLT-II Variable	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*	*95% CI for Exp (B)*	
							Lower	*Upper*
UN	6.07	1.73	12.37	1	.00**	431.68	14.67	12701.56
BN	-0.20	0.38	0.26	1	.61	0.82	0.39	1.74
PR	-1.58	0.54	8.55	1	.00*	0.21	0.07	0.59
BS	-0.11	0.36	0.10	1	.75	0.89	0.44	1.82
Constant	-0.71	0.42	2.90	1	.09	0.49		

Note. BN = semantically unrelated foils from list B; BS = semantically related foils from list B; CVLT-II = California Verbal Learning Test – Second Edition; MND = malingered neurocognitive dysfunction group; PR = novel and prototypically related foils; UN = novel and semantically unrelated foils.

*$p < .01$ **$p < .001$.

Examining Table 8 reveals that the likelihood that a person is putting forth suboptimal effort increases with the number of UN foils endorsed but decreases with the number of PR foils endorsed. More precisely, the odds that someone is exaggerating symptomatology increases by a factor of 431.68 for every UN foil endorsed but decreases by 4.76 for every additional PR foil endorsed, when all other factors are included in the model. Overall, this model suggests that suboptimal effort output is exemplified by a higher number of UN (novel-semantically unrelated) foils in light of fewer PR (novel-semantically related) foils. The foils consisting of list B items (i.e., BN and BS) do not significantly predict suboptimal effort output.

Using the Beta weights from each predictor making up the multivariate model, a logistic regression function was generated and exponentiated in order to calculate the probability of malingering for each case (see Equation 1; see Tabachnick & Fidell, 2001 for a review of working with logistic regression equations). Using this equation, the probabilities of group membership were calculated and saved for further analysis.

$$Pr\ (MND) = \frac{e^{-0.71 + 6.07(\text{UN raw score}) - 1.58(\text{PR raw score}) - 0.20(\text{BN raw score}) - 0.11(\text{BS raw score})}}{1 + e^{-0.71 + 6.07(\text{UN raw score}) - 1.58(\text{PR raw score}) - 0.20(\text{BN raw score}) - 0.11(\text{BS raw score})}} \quad (1)$$

Consistent with previous research (Ashendorf et al., 2003; Millis et al., 1995), separate univariate logistic regressions were performed for each CVLT-II predictor to assess their utility in differentiating between the TBI and MND groups in the event that these predictors were not included in the multivariate model. These analyses were also conducted to compare the full multivariate model with each partial univariate model. A test of the partial model with number of UN foils endorsed against the constant-only model was statistically significant, χ^2 (1, N = 113) = 47.96, p < .001, indicating that UN reliably distinguished between patients with quantifiable moderate-severe TBI and litigants putting forth suboptimal effort. This univariate model also differed significantly from the full model, χ^2 (3, N = 113) = 25.22, p < .001, suggesting that the full model accounted for more variance in MND status. Nonetheless, the partial model using UN as the sole predictor yielded good fit (Hosmer and Lemeshow Test p = .23) and showed improvement over the constant-only model ($-2LL_0$ = 132.78; $-2LL_{UN}$ = 84.82; Cox & Snell R^2 = .35; Nagelkerke R^2 = .50). Using the number of UN foils endorsed for classification yielded an overall hit rate of 83.2%; correctly identifying 64.5% of MND litigants and 90.2% of patients with moderate-severe TBI (PPP = 71.4%; NPP = 87.1%). Examining the odds ratio suggests that for every one unit increase in UN, the likelihood that an individual is exaggerating symptomatology increases by a factor of 8.13, all other factors excluded.

A test of the partial model with number of BN foils endorsed against the constant-only model was statistically significant, χ^2 (1, *N* = 113) = 9.58, *p* = .002, indicating that BN also reliably distinguished between patients with quantifiable moderate-severe TBI and litigants putting forth suboptimal effort. Comparing this model to the full model yielded a significant difference, χ^2 (3, *N* = 113) = 63.59, *p* < .001, suggesting that the full model accounted for more variance in MND status. This partial model showed good fit (Hosmer and Lemeshow Test *p* = .33) although the improvement over the constant-only model was limited ($-2LL_0$ = 132.78; $-2LL_{BN}$ = 123.20; Cox & Snell R^2 = .08; Nagelkerke R^2 = .12). Overall classification accuracy was 72.6% (specificity = 92.7%; sensitivity = 19.4%). Examining the odds ratio indicates that for every one unit increase in BN, the likelihood that an individual is exaggerating symptomatology increases by a factor of 1.50, when all other factors are excluded.

A partial model run with number of PR foils endorsed was not reliably different from the constant-only model, χ^2 (1, *N* = 113) = 0.25, *p* = .62, indicating that PR was not a reliable predictor of MND status on its own, corroborating that PR may have acted as a suppressor variable in the multivariate regression model. Finally, a test of the partial model using BS as a sole predictor was not significantly different from the constant-only model, χ^2 (1, *N* = 113) = 1.83, *p* = .18, suggesting it is not a useful predictor of MND status. Overall, the results suggest that the full model and the UN and BN univariate models reliably predict group membership, although BN’s classification accuracy is limited.

Logistic Regression Analyses for CVLT-II Composite Scores

A direct multiple logistic regression analysis was performed on MND status as outcome and two predictors based on the CVLT-II yes/no recognition composites: easy-to-reject items (ETR) and difficult-to-reject items (DTR). Preliminary analyses revealed adequate sample size-to-predictors ratio. A test of the full model with all the predictors against the constant-only model was statistically significant, χ^2 (2, N = 113) = 33.82, $p <$ .001, indicating that the predictors, as a set, reliably distinguished between patients with quantifiable moderate-severe TBI and litigants putting forth suboptimal effort. Based on the Hosmer and Lemeshow Goodness of Fit Test, the model had good fit (p = .41), indicating that model prediction did not differ significantly from observed values. Inspection of residuals revealed three outliers—all of which were participants from the MND group that were misclassified as belonging to the moderate-severe TBI group. These outliers were not removed because the model had good fit even with them included in the analysis. The model showed no problems with convergence and the standard errors for the parameters were not exceedingly large; thus, there was no multicollinearity evident (Tabachnick & Fidell, 2001). The full model showed moderate improvement over the constant-only model (-2LL_0 = 132.78; $-2\text{LL}_{\text{FULL2}}$ = 98.96; Cox & Snell R^2 = .26; Nagelkerke R^2 = .37). Using a .50 classification cut-off, the model correctly classified 81.4% of the participants, with 45.2% sensitivity and 95.1% specificity. The PPP was 77.8% and the NPP was 82.1%, suggesting very high predictive power (see Table 9). Leave-one-out cross-validation resulted in correct classification of 82.3% of the grouped cases, indicating excellent generalizability.

Table 9

Classification Table for Full Model Using the Composite Scores

		Predicted	
		TBI	MND
Observed	TBI	78	4
	MND	17	14

Note. MND = malingered neurocognitive dysfunction group; TBI = moderate-severe traumatic brain injury group.

Table 10 shows regression coefficients, Wald statistics, odds ratios, and 95% confidence intervals for the odds ratios for each predictor. According to the Wald criterion, the number of ETR foils and the number of DTR foils reliably predicted a person's effort output during testing. However, given that DTR emerged as a significant predictor in the multivariate regression model while having a near-zero (and not significant) correlation with the criterion variable suggests that it may have acted as a suppressor variable in the regression equation. Consistent with this contention is the fact that DTR's beta weight and its correlation with the criterion variable have different signs (Tabachnick & Fidell, 2001).

Table 10

Logistic Regression Analysis of MND Status as a Function of CVLT-II Composite Scores

CVLT-II Composite	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*	*95% CI for Exp (B)*	
							Lower	*Upper*
ETR	0.80	0.18	19.18	1	.00**	2.24	1.56	3.21
DTR	-0.32	0.13	6.42	1	.01	0.72	0.56	0.93
Constant	-1.29	0.38	11.64	1	.00*	0.28		

Note. CVLT-II = California Verbal Learning Test – Second Edition; DTR = difficult to reject foils; ETR = easy to reject foils; MND = malingered neurocognitive dysfunction group.

**p* < .01 ***p* < .001.

The likelihood that a person is putting forth suboptimal effort increases with the number of ETR foils endorsed, but decreases with the number of DTR foils endorsed. Stated differently, the likelihood that a test-taker is exaggerating or feigning symptoms increases by 2.24 times for every unit increase in ETR composite score but decreases by a factor of 1.39 for every additional DTR item endorsed. Overall, this model suggests that suboptimal effort output is characterized by an increased selection of ETR foils in light of fewer DTR items.

Based on the Beta weights from each predictor making up the multivariate model, a logistic regression function was generated and exponentiated in order to calculate the probability of malingering for each case (see Equation 2). Using this equation, the probabilities of group membership were calculated and saved for further analysis.

$$Pr\ (MND) = \frac{e^{-1.29 + 0.80(\text{ETR composite score}) - 0.32(\text{DTR composite score})}}{1 + e^{-1.29 + 0.80(\text{ETR composite score}) - 0.32(\text{DTR composite score})}} \qquad (2)$$

Consistent with the analyses conducted on the individual CVLT-II predictors, independent direct logistic regression analyses were conducted to compare the full multivariate model with each partial univariate model. A test of the partial model with ETR composite score against the constant-only model was statistically significant, χ^2 (1, N = 113) = 25.45, $p < .001$, indicating that ETR reliably distinguished between patients with TBI and litigants putting forth suboptimal effort. This partial model also differed significantly from the full model, χ^2 (1, N = 113) = 8.37, $p = .004$, suggesting that the full model accounted for more variance in MND status. The partial model using ETR composite score as the sole predictor yielded good fit (Hosmer and Lemeshow Test p = .19) and demonstrated a modest improvement over the constant-only model ($2LL_0$ = 132.78; $-2LL_{ETR}$ = 107.33; Cox & Snell R^2 = .20; Nagelkerke R^2 = .29). The ETR composite score yielded an overall hit rate of 79.6%; correctly identifying 51.6% of MND litigants and 90.2% of patients with moderate-severe TBI (PPP = 66.7%; NPP = 83.2%). Examining the odds ratio suggests that the likelihood that an individual is exaggerating symptomatology increases by a factor of 1.63 for every one unit increase in ETR composite score, all other factors excluded.

The partial model run with the DTR composite score was not reliably different from the constant-only model, χ^2 (1, N = 113) = 1.02, p = .31, indicating that DTR was not a reliable predictor of MND status on its own. Overall, the results suggest that the full model and the ETR composite reliably predict group membership.

Receiver Operating Characteristics (ROC) Curve

Model discriminability between the reference and the criterion group was very good for all variables as indicated by the area under the ROC curve (see Table 11).

Table 11

Area Under the ROC Curve for Significant Predictors of MND Status

Variable	*Area*	*SE*	*Asymptotic p*	*Asymptotic 95% CI*	
				Upper Bound	*Lower Bound*
Pr1	.92	.03	.00**	.88	.97
Pr2	.80	.05	.00**	.71	.89
UN	.80	.06	.00**	.69	.91
BN	.65	.06	.01	.53	.77
ETR	.73	.06	.00**	.61	.85

Note. BN = semantically unrelated foils from list B; ETR = easy to reject foils; Pr1 = predicted probability from equation 1; Pr2 = predicted probability from equation 2; ROC = Receiver operating characteristics; UN = novel and semantically unrelated foils.

*p < .01 **p < .001.

The best discriminability was obtained with the multivariate equation composed of the four CVLT-II foils (area under ROC curve = .92; p < .001), followed by the multivariate equation composed of the CVLT-II composite scores (area under ROC curve

= .80; *p* < .001) and by the individual UN foil (area under ROC curve = .80; *p* < .001; see Figure 1).

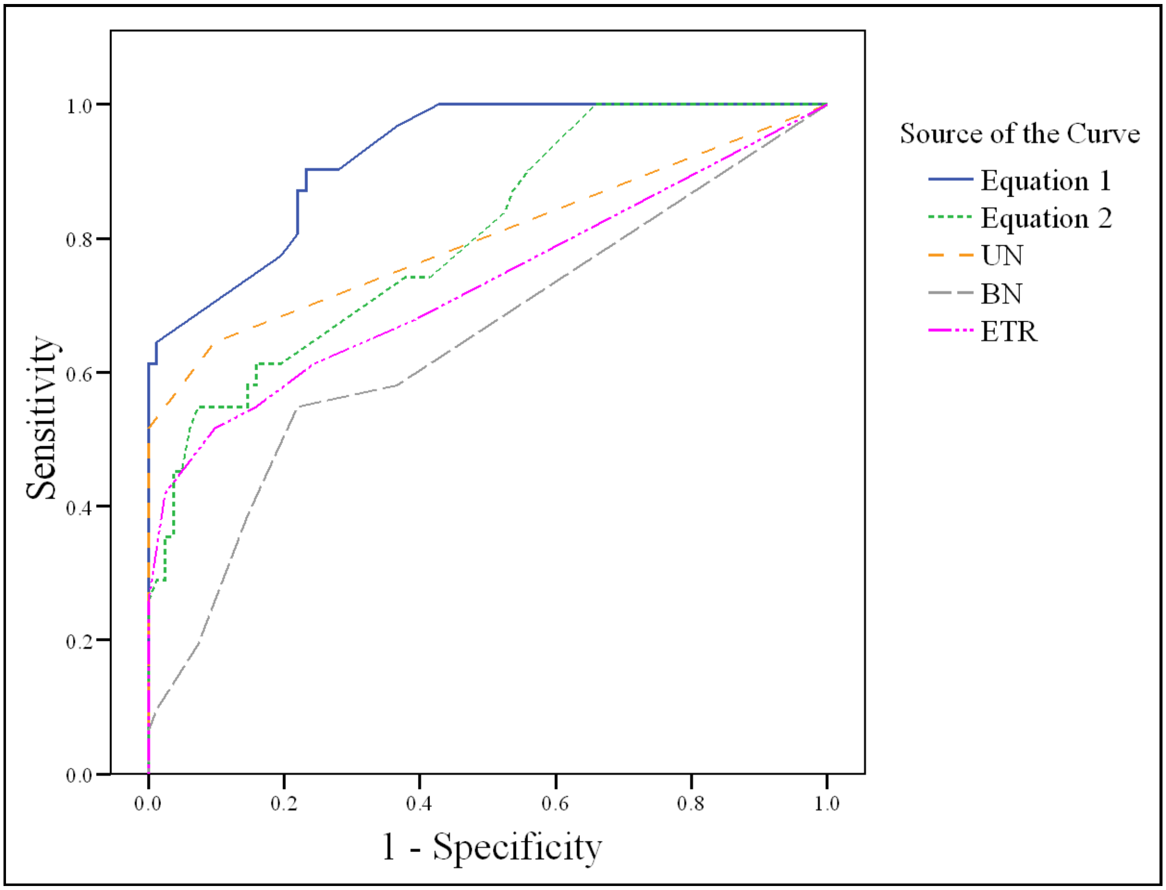

Figure 1. ROC curves for significant multivariate and univariate models predicting suboptimal effort.

Cut-off Scores Analyses

In an attempt to derive optimal cut-off scores, the ROC curves and frequency distributions of the various predictors were examined for the MND and moderate-severe TBI groups with the objective of selecting cut-off values that maintained the false positive error rate at a maximum of 10%. This emphasis on specificity minimized the likelihood that a patient with moderate-severe TBI would be classified as malingering.

Probability scores derived from both logistic equations were used for each multivariate model. Raw test scores were used for the UN, BN, and ETR variables. Table 12 displays the selected diagnostic cut-offs as well as the sensitivity, specificity, PPP, NPP, and overall hit rate associated with scores falling *on or above* each cut-off.

Table 12

Diagnostic Accuracy for Selected Predictors

Predictor	*Cut-off*	*Sensitivity*	*Specificity*	*Hit Rate*	*PPP*	*NPP*
Pr1	.38	64.5	98.8	89.4	95.3	88.0
	.77	61.3	100	87.6	100	85.4
Pr2	.44	54.8	92.7	82.3	73.9	84.5
	.48	51.6	93.9	82.3	76.1	83.7
	.50	45.2	95.1	81.4	77.7	82.1
	.54	41.9	96.3	81.4	81.0	81.5
	.71	35.5	97.6	80.6	84.8	80.0
	.79	29.0	98.8	79.7	90.1	78.7
	.82	25.8	100	79.6	100	78.1
UN	1	64.5	90.2	83.2	71.3	87.0
	2	51.6	100	86.7	100	84.5
BN	4	19.4	92.7	72.6	50.1	75.3
	5	9.7	98.8	74.3	75.3	74.3
	6	6.5	100	74.3	100	73.9
ETR	4	51.6	90.2	79.6	66.6	83.1
	5	41.9	97.6	82.3	86.8	81.6
	6	25.8	100	79.6	100	78.1

Note. All values are raw scores unless stated otherwise. BN = semantically unrelated foils from list B; ETR = easy to reject foils; NPP = Negative Predictive Power; PPP = Positive Predictive Power; Pr1 = predicted probability from equation 1; Pr2 = predicted probability from equation 2; UN = novel and semantically unrelated foils.

A cut-off score of .38 or higher for the probability scores generated by Equation 1 resulted in an overall correct classification rate of 89.4% with 98.8% of the moderate-severe TBI patients and 64.5% of the MND participants correctly classified. Given this study's base rate of 27.4%, participants scoring *at or above* .38 have a high likelihood that they are putting forth insufficient effort (i.e., PPP = 95.3%). Likewise, based on the NPP, there is an 88% chance that those scoring below this cut-off are giving adequate effort. Adjusting the cut-off value to .77 resulted in a slightly lower hit rate (87.6%), sensitivity (61.3%), and NPP (85.4%), but higher specificity (100%) and PPP (100%). Equation 2 also yielded promising diagnostic scores, but the cut-offs were higher and it displayed lower sensitivity than Equation 1, suggesting that Equation 1 is more precise at differentiating between moderate-severe TBI and MND samples. With respect to the univariate predictors, the UN variable appears to show diagnostic potential. Using a cut-off score of one resulted in a hit rate of 83.2%, with 64.5% sensitivity and 90.2% specificity. Adjusting the cut-off by one unit resulted in a correct classification of 86.7% cases (51.6% sensitivity and 100% specificity). Overall, using the abovementioned cut-off values for each predictor resulted in false positive error rates ranging between 0% and 9.8% depending on the index and the cut-off value used. All of these results fell within the 10% false positive error rate standard recommended by the extant literature (Greiffenstein et al., 1994; Millis, 1992), suggesting that these measures were good at minimizing false positive errors. Although the sensitivity scores were lower than the specificity scores, PPP values were very high for all measures, indicating that these embedded effort indices have considerable diagnostic utility, especially when using the more stringent cut-offs.

Because, as mentioned above, the base rates (prior probabilities) of incomplete effort vary considerably depending on the setting and population, PPP and NPP values were calculated for several hypothetical base rates using Bayes’ theorem as described by Millis (2003) and Millis and Volinsky (2001). Table 13 contains the PPP and NPP values for the UN, BN, and ETR variables as well as the two multivariate models for the five hypothetical base rates recommended by Greve and Bianchini (2004). Clinicians can use these cut-off values to assist their diagnosis of suboptimal effort output at varying degrees of stringency for various base rates depending on their setting.

Table 13

Diagnostic Accuracy of Predictors at Selected Cut-offs Varying in Stringency for Five Hypothetical Base Rates

		PPP					NPP				
Predictor	*Cut-off*	*BR= 10%*	*20%*	*30%*	*40%*	*50%*	*BR= 10%*	*20%*	*30%*	*40%*	*50%*
Pr1	.38	85.7	93.1	95.8	97.3	98.2	96.2	91.8	86.7	80.7	73.6
	.77	100	100	100	100	100	95.2	89.8	83.8	76.8	68.9
Pr2	.44	45.5	65.2	76.3	83.3	88.2	94.9	89.1	82.7	75.5	67.2
	.48	48.5	67.9	78.4	84.9	89.4	94.6	88.6	81.9	74.4	66.0
	.50	50.6	69.8	79.8	86.0	90.2	94.0	87.4	80.2	72.2	63.4
	.54	55.7	73.9	82.9	88.3	91.9	93.7	86.9	79.5	71.3	62.4
	.71	62.2	78.7	86.4	90.8	93.7	93.2	85.8	77.9	69.4	60.2
	.79	72.9	85.8	91.2	94.2	96.0	92.6	84.8	76.5	67.6	58.2
	.82	100	100	100	100	100	92.4	84.4	75.9	66.9	57.4
UN	1	42.2	62.2	73.8	81.4	86.8	95.8	91.0	85.6	79.2	71.8
	2	100	100	100	100	100	94.9	89.2	82.8	75.6	67.4
BN	4	22.8	39.9	53.2	63.9	72.7	91.2	82.1	72.9	63.3	53.5
	5	47.3	66.9	77.6	84.3	89.0	90.8	81.4	71.9	62.1	52.2
	6	100	100	100	100	100	90.6	81.1	71.4	61.6	51.7
ETR	4	36.9	56.8	69.3	77.8	84.0	94.4	88.2	81.3	73.7	65.1
	5	66.0	81.4	88.2	92.1	94.6	93.8	87.0	79.7	71.6	62.7
	6	100	100	100	100	100	92.4	84.4	75.9	66.9	57.4

Note. All values are raw scores unless stated otherwise. BN = semantically unrelated foils from list B; BR = base rate; ETR = easy to reject foils; NPP = Negative Predictive Power; PPP = Positive Predictive Power; Pr1 = predicted probability from equation 1; Pr2 = predicted probability from equation 2; UN = novel and semantically unrelated foils.

Examination of Misclassified TBI Cases

One individual from the moderate-severe TBI group was misclassified as putting forth insufficient or suboptimal effort using Equation 1. Closer examination of their responses showed that this case endorsed a considerable number of yes/no recognition false positive errors (i.e., 10 out of 32), indicating that they had low discriminability and were employing an affirmative response style. In addition, this individual performed in the impaired range with respect to the FCR task. A subsequent CIA showed that he performed worse than 100% of his age and education-matched peers, signifying that he missed many words in the FCR task that he correctly endorsed at least once during yes/no recognition, which is a rare finding in individuals putting forth adequate effort (Root et al., 2006). Overall, it is unclear whether this participant was putting forth adequate effort. The same participant was also misclassified as MND using the UN variable and ETR composite. That participant, however, was correctly classified by Equation 2. There was no discernible pattern found for the four TBI cases incorrectly classified as MND by Equation 2.

When examining the TBI cases that were incorrectly classified by the UN variable, it was evident that these seven cases committed significantly more recognition false positive errors (*M* = 10.38; *SD* = 2.87) than those TBI cases who were correctly classified as having a moderate-severe TBI (*M* = 3.84; *SD* = 2.87), $t(80) = -4.82$, $p < .001$. In addition, four out of the seven cases misclassified as MND showed impaired performance on the FCR task and a CIA of their responses demonstrated that they performed worse than 99.3% of their age and education-matched peers, suggesting insufficient effort output. Consistent with these findings, five out of seven misclassified cases received scores above .50 probability using Millis et al.'s (2007) formula. A similar pattern was observed when using the BN variable and ETR composite as predictors.

Examination of Misclassified MND Cases

With respect to false negative scores, although classification patterns differed between the predictors, three distinct groups could be formed. Group 1 consisted of 10 MND participants that were misclassified as putting forth adequate effort by all the predictors. Group 2 was composed of the six MND individuals that were correctly classified by all predictors. Finally, the remaining 15 cases made up Group 3 and varied in group membership (MND vs. TBI) depending on the predictor. A one-way between-subjects analysis of variance was conducted to explore the impact of group membership on recognition accuracy as determined by number of recognition hits, false positive errors committed, and recognition discriminability (*d'*). The means and standard deviations are presented in Table 14.

Table 14

Mean Raw Scores of Selected CVLT-II Recognition Variables for MND Sub-groups

	Group 1 (*n* =10)		Group 2 (*n* = 6)		Group 3 (*n* = 15)	
CVLT-II Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
False Positive Errors	0.90	0.88	15.83	3.54	8.87	3.76
Hits	11.10	3.41	10.33	2.34	10.67	2.32
d'	2.52	0.92	0.40	0.38	1.13	0.73

Note. All values are raw scores unless stated otherwise. CVLT-II = California Verbal Learning Test – Second Edition; *d'* = recognition discriminability; Group 1 = MND cases misclassified as TBI patients by all the CVLT-II predictors; Group 2 = MND cases correctly classified as MND; Group 3 = cases that varied in group classification depending on the predictor used.

There was a statistically significant difference in false positive errors committed for the three groups [$F(2, 28) = 46.04$, $p < .001$, partial $\eta^2 = .767$]. Post-hoc comparisons using the Bonferroni adjustment ($\alpha = .017$) indicated that the mean score for Group 1 ($M = 0.90$, $SD = 0.88$) was significantly different from Group 2 ($M = 15.83$, $SD = 3.54$) which was significantly different from Group 3 ($M = 8.87$, $SD = 3.76$). Likewise, there was a statistically significant difference in recognition discriminability (*d'*) for the three groups [$F(2, 28) = 17.46$, $p < .001$, partial $\eta^2 = .555$]. Post-hoc comparisons using the Bonferroni adjustment ($\alpha = .017$) indicated that the mean score for Group 1 ($M = 2.52$, $SD = 0.92$) was significantly better than the other two groups, but Group 2 ($M = 0.40$, *SD*

= 0.38) did not differ from Group 3 (*M* = 1.13, *SD* = 0.73). The three groups did not differ significantly in terms of Hits [*F*(2, 28) = 0.16, *p* = .85]. Overall, these findings suggest that the cases making up the MND group varied in terms of recognition accuracy. More precisely, MND cases committing fewer false positive recognition errors and displaying better recognition discriminability were more likely to be wrongly classified as belonging to the moderate-severe TBI group. Although these results are limited due to the small and uneven sample sizes, they do suggest that the total number of false positive errors in the yes/no recognition trial of the CVLT-II may have potential as a predictor of suboptimal effort output.

Consequently, a direct logistic regression analysis was performed on MND status as outcome and total false positive errors committed in the yes/no recognition trial of the CVLT-II. The partial model with one predictor against the constant-only model was statistically significant, χ^2 (1, *N* = 113) = 9.00, *p* = .003, indicating that false positive errors reliably distinguished between patients with quantifiable moderate-severe TBI and litigants putting forth suboptimal effort. Based on the Hosmer and Lemeshow Goodness of Fit Test, the model had good fit (*p* = .05), indicating that model prediction did not differ significantly from observed values. Using a .50 classification cut-off, the model correctly classified 75.2% of the participants, with 16.1% sensitivity and 97.6% specificity (PPP = 75.5%; NPP = 81.4%). Examination of the frequency distribution indicated that a score above 11 false positive errors results in 91.5% specificity and 32.3% sensitivity, whereas a score above 13 results in the classification accuracy determined by the logistic regression (i.e., 16.1% sensitivity and 97.6% specificity), and a score above 15 results in 100% specificity to the detriment of very low sensitivity (i.e., 6.5%).

Post-hoc Ratio Analyses

Following from the results above, several post-hoc ratio analyses were conducted. Specifically, given that the likelihood that a person is putting forth suboptimal effort increases with the number of UN foils endorsed and decreases with the number of PR foils endorsed (see Table 8), a UN/PR ratio was generated. In addition, given that UN had positive Beta weights and the rest of the foils had negative Beta weights on Equation 1, a separate ratio – UN/(PR+BN+BS) – was generated. Likewise, an ETR/DTR ratio was generated in order determine the effect of a semantically-driven pattern of responses on MND status. Of note, because some individual predictors had values of zero (i.e., cannot be used as denominator), the total sample size of these ratios was smaller than that of the original sample. Following these ratio calculations, separate univariate logistic regression analyses were performed on MND status as outcome and the three ratios as predictors.

A test of the UN/PR model against the constant-only model was statistically significant, χ^2 (1, N = 86) = 63.60, $p < .001$, indicating that UN/PR reliably distinguished between patients with moderate-severe TBI and litigants putting forth suboptimal effort. This model showed good fit (Hosmer and Lemeshow Test $p = .76$) and good improvement over the constant-only model ($-2LL_0 = 99.88$; $-2LL_{UN/PR} = 36.28$; Cox & Snell $R^2 = .52$; Nagelkerke $R^2 = .76$). As shown on Table 15, a cut-off of .50 yielded a classification accuracy of 95.3% (specificity = 98.4%; sensitivity = 87.0%; PPP = 95.2%; NPP = 95.4%; base rate = 26.7%).

Table 15

Classification Table for the UN/PR Model

		Predicted	
		TBI	MND
Observed	TBI (n = 63)	62	1
	MND (n = 23)	3	20

Note. MND = malingered neurocognitive dysfunction group; PR = novel, semantically related foils; TBI = moderate-severe traumatic brain injury group; UN = novel, semantically unrelated foils.

A test of the model with UN/(PR+BN+BS) against the constant-only model was statistically significant, χ^2 (1, N = 94) = 49.16, $p < .001$, indicating that this predictor also reliably distinguished between patients with moderate-severe TBI and litigants putting forth suboptimal effort. This model showed good fit (Hosmer and Lemeshow Test p = .32) and improvement over the constant-only model ($-2LL_0 = 112.74$; $-2LL_{UN/(PR+BN+BS)} = 63.58$; Cox & Snell R^2 = .41; Nagelkerke R^2 = .58). Using a classification cut-off of .50, overall classification accuracy was 89.4% (specificity = 97.0%; sensitivity = 70.4%; PPP = 90.5%; NPP = 89.0%; base rate = 28.7%; see Table 16).

Table 16

Classification Table for the UN/(PR+BN+BS) Model

		Predicted	
		TBI	MND
Observed	TBI (n = 67)	65	2
	MND (n = 27)	8	19

Note. BN = semantically unrelated foils from list B; BS = semantically related foils from list B; MND = malingered neurocognitive dysfunction group; PR = novel, semantically related foils; TBI = moderate-severe traumatic brain injury group; UN = novel, semantically unrelated foils.

Finally, a test of ETR/DTR model against the constant-only model was statistically significant, χ^2 (1, N = 92) = 13.22, $p < .001$, indicating that ETR/DTR also reliably distinguished between patients with moderate-severe TBI and litigants putting forth suboptimal effort. This model showed good fit (Hosmer and Lemeshow Test p = .07) and adequate improvement over the constant-only model ($-2LL_0$ = 109.55; $-2LL_{ETR/DTR}$ = 96.33; Cox & Snell R^2 = .13; Nagelkerke R^2 = .19). Using a classification cut-off of .50, the model correctly classified 79.3% of participants (specificity = 95.5%; sensitivity = 38.5%; PPP = 76.9%; NPP = 79.7%; base rate = 28.3%; see Table 17).

Table 17

Classification Table for the ETR/DTR Model

		Predicted	
		TBI	MND
Observed	TBI (*n* = 66)	63	3
	MND (*n* = 26)	16	10

Note. DTR = difficult to reject foils; ETR = easy to reject foils; MND = malingered neurocognitive dysfunction group; TBI = moderate-severe traumatic brain injury group.

ROC analyses were conducted in order to evaluate model discriminability and generate cut-off scores. The best discriminability was obtained with the UN/PR ratio (area under ROC curve = .92; $p < .001$), followed by the UN/(PR+BN+BS) ratio (area under ROC curve = .85; $p < .001$) and the ETR/DTR ratio (area under ROC curve = .77; $p < .001$; see Figure 2). Table 18 displays the diagnostic cut-offs for each predictor as well as the sensitivity, specificity, PPP, NPP, and overall hit rate associated with scores falling on or above each cut-off. As shown on Table 18, a cut-off score of .23 or higher for the UN/PR ratio resulted in a classification accuracy of 91.9%, with 93.7% of moderate-severe TBI patients and 87.0% of MND participants correctly classified. Raising the cut-off value to .30 or higher resulted in improved specificity (i.e., 98.4%) but no change in sensitivity. With the UN/PR analysis' base rate of 26.7%, the PPP values ranged from 83.5% to 100% while the NPP values ranged from 88.7% to 95.2%

depending on the cut-off values used. With respect to the UN/(PR+BN+BS) ratio, the best classification rate (i.e., 90.4%) was achieved with a cut-off of .15 or higher, resulting in 70.4% sensitivity and 98.5% specificity. With a base rate of 28.7%, the PPP values ranged from 76.8% to 100% while the NPP values ranged from 79.8% to 89.7% depending on the cut-off used. Finally, the ETR/DTR ratio resulted in a hit rate of 83.7%, with 61.5% sensitivity and 92.4% specificity when the cut-off was .59 or higher. At a base rate of 28.3%, the PPP varied from 60.7% to 81.7% while the NPP varied from 73.1% to 85.9%. Overall, these ratios show promise in distinguishing between moderate-severe TBI and MND status.

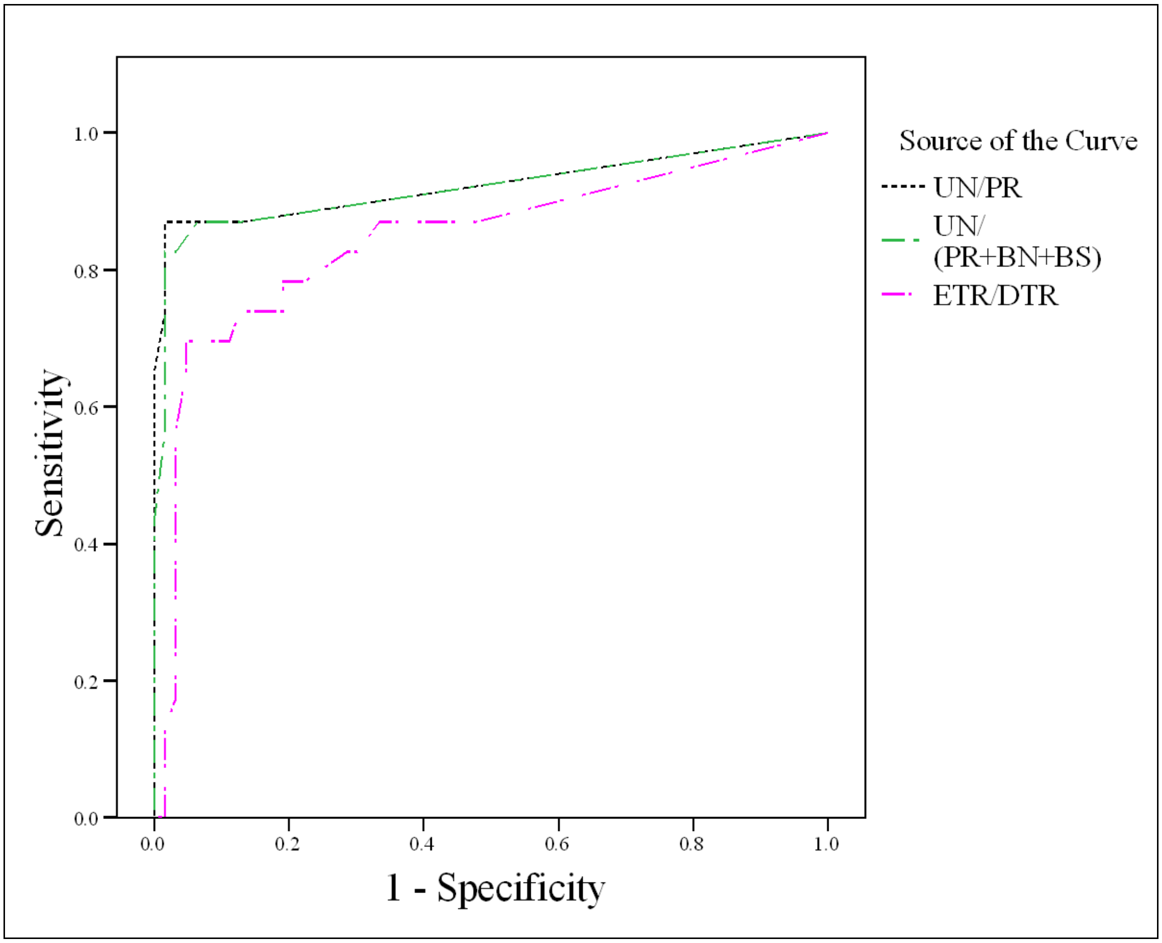

Figure 2. ROC curves for significant CVLT-II ratio models predicting suboptimal effort.

Table 18

Diagnostic Accuracy Statistics for Ratios

Predictor	*Cut-off*	*Sensitivity*	*Specificity*	*Hit Rate*	*PPP*	*NPP*
UN/PR	.23	87.0	93.7	91.9	83.5	95.2
	.30	87.0	98.4	95.4	95.2	95.4
	.57	65.2	100.0	90.7	100.0	88.7
UN/(PR+BN+BS)	.09	74.1	91.0	86.1	76.8	89.7
	.15	70.4	98.5	90.4	95.0	89.2
	.27	37.0	100.0	81.9	100.0	79.8
ETR/DTR	.59	61.5	92.4	83.7	76.1	85.9
	.68	50.0	95.5	82.4	81.7	82.6
	1.43	11.5	97.0	72.4	60.7	73.1

Note. All values are raw scores unless stated otherwise. BN = semantically unrelated foils from list B; BS = semantically related foils from list B; DTR = difficult to reject foils; ETR = easy to reject foils; NPP = Negative Predictive Power; PPP = Positive Predictive Power; PR = novel, semantically related foils; UN = novel, semantically unrelated foils.

Analyses of MVA Subsample

Because the original sample consisted of heterogeneous mechanisms of injury, the analyses performed above were repeated using only participants that were involved in MVAs. These extra analyses were conducted in order to investigate whether the CVLT-II predictors would continue to differentiate between MND and moderate-severe TBI using *relatively* more homogeneous reference and criterion groups. The moderate-severe TBI subgroup (sTBI) consisted of 25 patients who were in MVAs, 6 patients who were pedestrians in MVAs, and 4 patients who were on motorcycles when an MVA took place. The MND subgroup (sMND) consisted of 24 participants in MVAs, 4 pedestrians in MVAs, and 1 person who was on a motorcycle when an MVA occurred. Because these groups were derived from the larger reference and criterion groups, the inclusion criteria were identical to those mentioned above (see Table 19 for descriptive statistics).

Table 19

Descriptive Statistics for the Participants Making up the Reference and Criterion Subgroups

	sMND (n =29)		sTBI (n = 35)		T-Test		
Variable	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
Age	36.80	10.99	43.00	12.28	-2.13	62	.04
Years of education	12.77	2.30	12.79	2.65	-0.35	62	.97
Months post injury	69.03	52.52	53.69	48.57	1.20	62	.23

Note. sMND = malingered neurocognitive dysfunction MVA subgroup; sTBI = moderate-severe traumatic brain injury MVA subgroup.

A direct multiple logistic regression analysis was performed on sMND status as outcome and four predictors based on the CVLT-II yes/no recognition foils. A test of the full model with all the predictors against the constant-only model yielded statistically significant results, χ^2 (4, N = 64) = 47.64, $p < .001$, indicating that the predictors, as a set, reliably distinguished between patients with moderate-severe TBI and litigants putting forth suboptimal effort. Based on the Hosmer and Lemeshow Goodness of Fit Test, the model had good fit (p = .72), indicating that model prediction did not differ significantly from observed values. The full model evidenced good improvement over the constant-only model ($-2LL_0 = 88.16$; $-2LL_{FSUB1} = 40.52$; Cox & Snell R^2 = .53; Nagelkerke R^2 = .70). Using a .50 classification cut-off, the model correctly classified 85.9% of the participants, with 72.4% sensitivity and 97.1% specificity. The PPP was 95.5% and the NPP was 80.9%, indicating very high predictive power.

A separate direct multiple logistic regression analysis was performed on sMND status as outcome and two composites based on the CVLT-II yes/no recognition foils. A test of the full model with both predictors against the constant-only model yielded statistically significant results, χ^2 (2, N = 64) = 23.62, $p < .001$, indicating that the predictors, as a set, reliably distinguished between patients who sustained a moderate-severe TBI following an MVA and litigants putting forth suboptimal effort after an MVA. Based on the Hosmer and Lemeshow Goodness of Fit Test, the model had good fit (p = .27), indicating that model prediction did not differ significantly from observed values. The full model evidenced good improvement over the constant-only model ($-2LL_0 = 88.16$; $-2LL_{FSUB2} = 64.54$; Cox & Snell $R^2 = .31$; Nagelkerke $R^2 = .41$). Using a .50 classification cut-off, the model correctly classified 73.4% of the participants, with 55.2% sensitivity and 88.6% specificity. The PPP was 80.0% and the NPP was 70.4%, denoting very high predictive power. Using the Beta scores from each predictor making up the multivariate models, two equations were generated to calculate the probability of malingering for each case making up the subgroups (see Equation 3 and 4).

$$Pr\ (sMND) = \frac{e^{-0.64 + 5.91(\text{UN raw score}) - 1.53(\text{PR raw score}) - 0.45(\text{BN raw score}) + 0.06(\text{BS raw score})}}{1 + e^{-0.64 + 5.91(\text{UN raw score}) - 1.53(\text{PR raw score}) - 0.45(\text{BN raw score}) + 0.06(\text{BS raw score})}} \quad (3)$$

$$Pr\ (sMND) = \frac{e^{-0.52 + 0.83(\text{ETR composite score}) - 0.32(\text{DTR composite score})}}{1 + e^{-0.52 + 0.83(\text{ETR composite score}) - 0.32(\text{DTR composite score})}} \quad (4)$$

Using the aforementioned equations, probabilities of subgroup membership were calculated and entered into an ROC analysis, along with the variables found to load significantly into the models, in order to determine model discriminability and cut-off scores. The best discriminability was obtained with the multivariate equation composed of the four CVLT-II foils (area under ROC curve = .94; $p < .001$), followed by the multivariate equation composed of the CVLT-II composite scores (area under ROC curve

= .80; *p* < .001), the individual UN foil (area under ROC curve = .80; *p* < .001), the ETR composite score (area under ROC curve = .74; *p* = .001), and the individual BN foil (area under ROC curve = .66; *p* = .031; see Figure 3). Consistent with the procedures employed on the larger sample, cut-off values were determined by examining the ROC curves and frequency distributions for the various predictors.

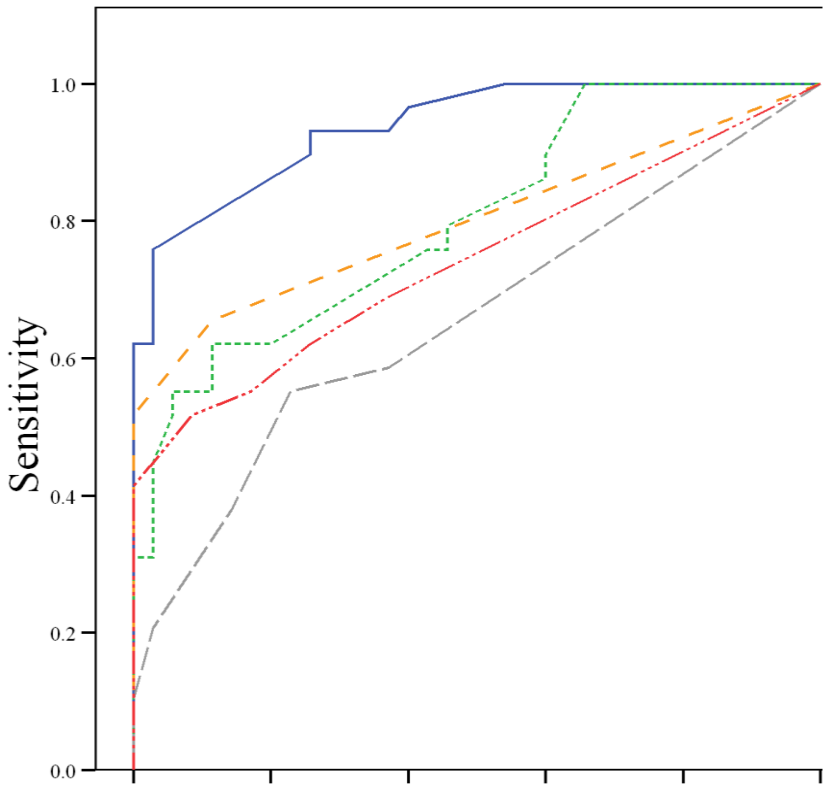

Figure 3. ROC curves for significant multivariate and univariate models predicting suboptimal effort in the MVA subgroup.

Table 20 displays the diagnostic cut-offs for each significant predictor as well as the sensitivity, specificity, PPP, NPP, and overall hit rate associated with scores falling on or above each cut-off.

Table 20

Diagnostic Accuracy for Selected Predictors Used in the MVA Subsample

Predictor	*Cut-off*	*Sensitivity*	*Specificity*	*Hit Rate*	*PPP*	*NPP*
Pr3	.38	89.7	74.3	81.3	74.3	89.7
	.49	75.9	97.1	87.5	95.6	82.9
	.86	62.1	100	82.8	100	76.1
Pr4	.47	62.1	88.6	76.6	81.9	73.8
	.49	58.6	88.6	75.0	81.0	72.1
	.66	55.2	94.3	76.6	88.9	71.8
	.69	51.7	94.3	75.0	88.3	70.2
	.72	44.8	97.1	73.4	92.8	68.0
	.75	37.9	97.1	70.3	91.5	65.4
	.88	31.0	100	68.7	100	63.6
UN	1	65.5	88.6	78.1	82.6	75.6
	2	51.7	100	78.1	100	71.4
BN	4	20.7	97.1	62.5	85.5	59.6
	5	10.3	100	59.4	100	57.4
ETR	4	51.7	91.4	73.4	83.3	69.5
	5	41.4	100	73.4	100	67.3

Note. All values are raw scores unless stated otherwise. BN = semantically unrelated foils from list B; ETR = easy to reject foils; MVA = motor vehicle accident; NPP = Negative Predictive Power; PPP = Positive Predictive Power; Pr3 = predicted probability from equation 3; Pr4 = predicted probability from equation 4; UN = novel and semantically unrelated foils.

The probability values generated from Equations 3 and 4 consistently yielded good sensitivity and specificity across different cut-off scores, as did the univariate predictors and ETR composite score. Given this sample's MND base rate of 45.3%, the classification rates and predictive power values derived by using the subgroups were similar across predictors and cut-off values to those attained using the original sample, especially when comparing said values to those on Table 13 under base rates of 40% and 50%. Thus, all the predictors distinguished between MND status and moderate-severe TBI when using a subsample of MVA-only cases, albeit with varying degrees of certainty. For example, at a base rate of 45.3%, litigants who have been in an MVA and score at or above .49 on Equation 3 have a very high likelihood that they are putting forth suboptimal effort (i.e., PPP = 95.6%). Similarly, endorsing one or more UN foils is associated with an 82.6% chance that the participant is giving poor effort. As with other predictors and samples in this study, manipulating the cut-off values can help the clinician to adjust the diagnostic accuracy to their needs.

Cross-Validation of Original Sample with Other Models

As noted in the multivariate analyses of the original sample, validation using a leave-one-out procedure resulted in excellent cross-classification for both models (i.e., 86.7% and 82.3% respectively). Performing cross-validation analyses with the leave-one-out procedure for each univariate predictor resulted in cross-classification values of 86.7%, 72.6%, and 79.6% for the UN, BN, and ETR predictors, respectively. In order to generalize the findings of this study to other samples, several direct logistic regression analyses were performed with CVLT and CVLT-II variables previously shown to predict MND status in different samples. Table 21 displays the predictors used along with their sources.

Table 21

Predictors Used in Cross-Validation

Predictors Used	Source
CVLT Total 1-5, Hits, LDCR, and *d'*	Millis et al., 1995; Sweet et al., 2000
CVLT Hits	Coleman et al., 1998; Curtis et al., 2006
CVLT-II LDFR, *d'*, and Recall Discriminability SS	Millis et al., 2007

Note. CVLT = California Verbal Learning Test; CVLT-II = California Verbal Learning Test – Second Edition; *d'* = recognition discriminability; Hits = recognition hits; LDCR = long-delay cued recall; LDFR = long-delay free recall; SS = scaled score; Total 1-5 = total number of words learned across five trials.

A test of the full model with all the CVLT predictors from Millis et al.’s (1995) study against the constant-only model was statistically significant, χ^2 (4, N = 113) = 29.14, $p < .001$, indicating that the predictors, as a set, reliably distinguished between participants with moderate-severe TBI and litigants putting forth suboptimal effort. According to the Wald criterion, only number of hits reliably predicted MND status, z = 6.19, $p = .01$, $B = -0.352$, indicating that MND status was characterized by fewer recognition hits. The model correctly classified 78.8% of the cases, with 54.8% sensitivity and 87.8% specificity. These values were different from those reported by Millis et al. (1995; 83% sensitivity and 96% specificity) but similar to those reported by Sweet et al. (2000). The generalizability of CVLT hits as a predictor of MND status was assessed by examining the ROC curve and frequency distributions at the cut-offs previously reported (Coleman et al., 1998; Curtis et al., 2006). Applying the cut-off of 11 to the present sample correctly classified 48.4% of the MND cases and 90.2% of the TBI cases (Area under the ROC curve = .79, $p < .001$), which are comparable to those reported by Curtis and collaborators (2006; 47% sensitivity and 96% specificity).

Finally, a direct logistic regression analysis was performed on MND status as outcome and the three BMA-derived CVLT-2 predictors from Millis et al.’s (2007) study: long-delay free recall, recall discriminability standard score, and recognition discriminability. A test of the full model against the constant-only model was statistically significant, χ^2 (3, N = 113) = 23.56, $p < .001$, indicating that the predictors, as a set, reliably distinguished between patients with moderate-severe TBI and litigants putting forth suboptimal effort. The Hosmer and Lemeshow Test indicated the presence of good fit ($p = .28$). The full model evidenced moderate improvement over the constant-only model ($2LL_0$ = 132.78; $-2LL_{BMA}$ = 109.22; Cox & Snell R^2 = .19; Nagelkerke R^2 = .27). Using a .49 classification cut-off, the model correctly classified 81.4% of the participants, with 51.6% sensitivity and 92.7% specificity. With a cut-off of .45, the model correctly classified 78.8% of the participants, with 54.8% sensitivity and 87.8% specificity. These classification values were somewhat better than those values reported by Millis et al. (2007), which were 68% sensitivity and 84% specificity using the first cut-off and 73% sensitivity and 81% specificity using the second cut-off. The apparent discrepancy between the present study's classification accuracy and Millis et al.'s (2007) findings could be partially explained by the fact that the inclusion criteria for their litigant group consisted of having failed only one effort measure, whereas the inclusion criteria herein require each MND participant to fail at least two measures.

Cross-Validation of Predictors with Other Samples

To validate the predictors further, the multivariate and univariate cut-offs determined above were used to classify two additional samples retrieved from RIM—patients with complicated mild TBI supplying adequate effort (MTBI), and patients with complicated mild TBI performing poorly on effort measures (SE). Table 22 summarizes the most salient descriptive data making up the criterion and reference groups.

Table 22

Descriptive Statistics of the Participants Used in Cross-Validation

	MND (*n* =31)		MTBI (*n* = 19)		SE (*n* = 23)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	42.29	12.39	48.11	18.60	51.96[a]	15.03
Years of education	12.65	2.65	13.11	2.83	12.70	2.76
Months post injury	56.39	49.43	34.94	38.47	31.09	39.59

Note. MND = malingered neurocognitive dysfunction group; MTBI = patients with complicated mild TBI supplying adequate effort; SE = patients with complicated mild TBI performing poorly on effort measures.

[a] Differed from the MND group on this parameter, $p < .05$.

The MTBI group consisted of 19 outpatients (12 male, 7 female) between the ages of 23 and 75 (M = 48.11; SD =18.60) and ranging in education from 9 to 18 years (M = 13.11; SD = 2.83). An independent samples t-test conducted revealed no significant differences between the MTBI and MND groups for age, $t(48) = 1.33$, $p = .19$, education, $t(48) = 0.58$, $p = .56$, or months post-injury, $t(47) = -1.58$, $p = .12$. In addition, chi-square analyses revealed that the two groups did not differ in regards to gender proportions, χ^2 (1, N = 50) = 0.00, $p = 1.00$ (Yates' Correction for Continuity), or in proportion of different ethnicities, χ^2 (3, N = 50) = 2.53, $p = .47$.

The SE group consisted of 23 outpatients (15 male, 8 female) between the ages of 48 and 63 (M = 51.96; SD =15.03) and ranging in education from 7 to 18 years (M = 12.70; SD = 2.76). An independent samples t-test revealed that the SE group was significantly older than the MND group, $t(52) = 2.59$, $p = .01$. No significant differences were found between the SE and MND groups for education, $t(52) = 0.07$, $p = .95$, or months post-injury, $t(51) = -1.99$, $p = .052$, although the latter approached significance. In addition, chi-square analyses revealed that the two groups did not differ in regards to gender proportions, χ^2 (1, N = 54) = 0.00, $p = 1.00$ (Yates' Correction for Continuity), or in proportion of different ethnicities, χ^2 (3, N =54) = 3.25, $p = .35$.

Both groups consisted of patients who had experienced brief (i.e., < 30 minutes) or no LOC or PTA, and who had scored at or above 13 on the GCS. However, unlike the individuals in the MND group, all the patients making up the MTBI and SE groups had visible pathology on neuroimaging or showed positive neurological signs. In addition, there were no cases in active litigation or pursuing worker's compensation at the time of testing. Thus, the MTBI and SE groups consisted of patients with complicated mild TBIs who were undergoing a neuropsychological evaluation and who presented with no identifiable external incentive. The difference between the MTBI and SE groups, however, was that the former failed fewer than two effort measures, whereas the latter failed more than two.

Specificity values were determined by examining the ROC curves and frequency distributions at various cut-offs as derived above. Probability scores derived from both logistic equations were used to represent each multivariate model. Raw test scores were used for the UN, BN, and ETR variables. Table 23 displays the specificity values observed at or above selected diagnostic cut-offs for the MTBI and SE groups. Equation 1, the UN variable, and the ETR composite score required a higher cut-off value in order maintain the false positive error rate at a maximum of 10% for the highest cut-off value. Overall, taking into account the minor adjustment in cut-off values, the predictors derived from the initial analyses perform relatively well when classifying patients with complicated mild TBI putting forth adequate effort (i.e., MTBI). On the other hand, the classification rates were unimpressive with respect to patients with complicated mild TBI who performed poorly on effort tests (i.e., SE). For this group, the predictors failed to reach the minimum standard specificity value of 90% on most selected cut-offs. This finding indicates that the cut-off values might need to be raised when differentiating between litigants putting forth suboptimal effort and patients with complicated mild TBIs performing poorly on effort tests. It is important to note, however, that the SE group was significantly older than the MND group, which might have influenced the resulting classification accuracy, especially given that age has been reported to have a moderate negative effect on CVLT-II performance (Delis et al., 2000). Thus, because older age would be expected to reduce the performance by the SE group, the specificity values reported are likely underestimates of actual values. Overall, the multivariate and univariate models appear to differentiate adequately between different samples varying in brain injury severity and effort output when the cut-offs are adjusted accordingly.

Table 23

Predictive Accuracy for the MTBI and SE Groups for Selected Predictors

Predictor	*Cut-off*	*(MND) Sensitivity*	*MTBI Specificity*	*PPP*	*NPP*	*SE Specificity*	*PPP*	*NPP*
Pr1	.41	64.5	84.2	86.0	78.1	82.6	83.3	75.8
	.69	61.3	89.5	89.8	77.6	82.6	82.6	74.2
	.96[a]	51.6	100	100.0	75.6	91.3	88.9	71.8
Pr2	.48	51.6	89.5	88.1	73.5	82.6	80.0	69.7
	.54	41.9	89.5	85.7	69.8	82.6	76.4	65.7
	.67	35.5	94.7	90.9	68.8	87.0	78.6	64.5
	.77	29.0	94.7	89.1	66.7	91.3	81.8	63.4
	.80	29.0	100	100.0	67.9	95.7	90.1	64.5
UN	1	64.5	73.7	78.6	75.7	60.9	69.0	69.8
	2	51.6	94.7	93.6	74.6	82.6	80.0	69.7
	3[a]	19.4	100	100.0	65.0	91.3	75.0	60.4
BN	4	19.4	94.7	84.6	63.8	87.0	66.8	59.3
	5	9.7	100	100.0	62.4	91.3	60.0	57.7
	6	6.5	100	100.0	61.6	95.7	67.1	58.0
ETR	4	51.6	89.5	88.1	73.5	78.3	76.2	68.6
	5	41.9	94.7	92.2	71.0	82.6	76.4	65.7
	6	25.8	100	100.0	66.9	87.0	72.8	61.2
	7[a]	16.1	100	100.0	64.1	87.0	62.5	58.3
	8[a]	12.9	100	100.0	63.3	95.7	80.2	59.7

Note. All values are raw scores unless stated otherwise. [a]Additional cut-off values were incorporated to achieve higher specificity. BN = semantically unrelated foils from list B; ETR = easy to reject foils; MTBI = patients with complicated mild TBI giving good effort; Pr1 = predicted probability from equation 1; Pr2 = predicted probability from equation 2; SE = patients with complicated mild TBI performing poorly on effort measures; UN = novel and semantically unrelated foils.

Discussion

The present study sought to determine whether different types of CVLT-II yes/no recognition foils and composite scores could reliably distinguish between a group of participants with quantifiable moderate-severe traumatic brain injuries and a group of individuals with mild or no brain injuries who were actively pursuing compensation and who had shown suboptimal effort output on at least two stand-alone tests or indices from psychometric tests. The moderate-severe TBI group had objectively documented neuropathology and unequivocal evidence of traumatic brain injuries as per several acute indices of injury severity. The individuals in the MND group had sustained questionable, if any, mild head injuries and were expected to have no prolonged effects based on the time elapsed between injury and assessment, but they complained of chronic and pervasive cognitive difficulties affecting day-to-day functioning and displayed abnormally poor performance on neuropsychological tests. Although none of the individuals making up the MND group honestly divulged exaggerating their symptoms, based on their performance on SVTs and neuropsychological indices of effort it was reasonable to characterize them as supplying suboptimal effort or feigning symptoms. Given findings from previous outcome studies (e.g., Dikmen et al., 1995; Rohling et al., 2003; Schretlen & Shapiro, 2003), a dose-response relationship between injury severity and cognitive impairment was expected whereby the group with mild TBI should outperform the group with moderate-severe TBI on neuropsychological tests. Thus, any significant differences in performance between the groups in the unexpected direction (mild TBI worse than moderate-severe TBI) would be considered inconsistent with expected patterns of brain functioning, implicating alternative reasons for the discrepancy.

Consistent with previous research in the area of malingering, this study sought to formulate an embedded effort measure that was so straightforward that even those with severe TBI could perform well. Along these lines, impaired performance by relatively intact individuals (i.e., mild or no TBI) would be considered aberrant compared to a moderate-severe TBI population, thus making it suspicious in terms of symptom exaggeration or suboptimal effort output. That was the case in employing the yes/no recognition trial of the CVLT-II. During the administration of the CVLT-II, by the time the individual is asked to recognize the target words from a list of foils, they have been exposed to the target items 5 times during the acquisition phase and have been required to recall them twice after a short delay and twice after a longer delay. Thus, irrespective of how many correct targets they endorse, it would be expected that individuals endorse few recognition foils (i.e., false positive errors) unless other factors are involved that prevent the original encoding of target items or impede their retrieval. In individuals with different types of neurological conditions causing memory impairment, either or both scenarios may be involved depending on the type, location, and severity of their neuropathology. In the case of uninjured responders, however, these types of errors are likely due to other factors outside of brain injury, including psychological distress, fluctuations in attention and/or motivation, or negative response bias, the latter of which would be suspected in the presence of other contextual factors such as secondary gain (e.g., Binder & Willis, 1991; Millis & Volinsky, 2001).

The findings of this study suggest that there is considerable merit in using the different types of false positive recognition errors in the CVLT-II to detect suboptimal effort output. As hypothesized, the moderate-severe TBI group endorsed fewer UN foils, BN foils, and ETR items than the MND group. In addition, the group of foils and composites as separate probability equations resulted in excellent discriminability between the reference and criterion groups and proved better at predicting suboptimal effort than the individual foils or composites. Examining Tables 12, 20, and 23 reveals that these predictors provided very good discrimination between litigants with questionable head injuries putting forth suboptimal effort and non-litigating patients with complicated mild and moderate-severe TBI without incentive for negative response bias. More specifically, the UN variable yielded very high classification rates depending on the cut-off used. A raw score of 1 or higher resulted in the correct classification of 64.5% of malingerers and 90.2% of patients with moderate-severe TBI. Using a more stringent cut-off of greater than or equal to 2 resulted in 51.6% sensitivity, 100% specificity, and an overall hit rate of 86.7%. This cut-off value was also effective at correctly excluding 94.7% of patients with complicated mild TBI putting forth adequate effort and 82.6% of patients with complicated mild TBI who failed at least two effort measures (see Table 23). Overall, the UN variable appears to differentiate adequately between litigants who are malingering and patients with traumatic head injuries ranging from complicated mild to moderate-severe. With respect to other individual predictors, the BN variable showed some, albeit limited, utility at detecting suboptimal effort but had adequate specificity for moderate-severe TBI as well as complicated mild TBI, especially when using a cut-off score of 5 or higher. Finally, the diagnostic accuracy of the ETR variable was also very good albeit slightly lower than that of the UN variable. ETR scores falling on or above 5 detected 13 out of 31 individuals from the MND group (41.9% sensitivity) while correctly rejecting 80 out of 82 individuals with moderate-severe TBI (97.6% specificity), 18 out of 19 patients with complicated mild TBI and good effort output (94.7% specificity), and 19 out of 23 patients with complicated mild TBI who failed two or more effort indices (82.6% specificity), the latter of which falls below suggested guidelines. Using a more stringent cut-off value of 6 or higher resulted in lower overall classification rates but specificity values approaching 100%, which resulted in higher probabilities that those scoring above these cut-offs were actually putting forth suboptimal effort.

Multivariate equations using CVLT and CVLT-2 variables have been previously found effective at differentiating between malingerers and patients with head injuries (e.g., Bauer et al., 2005; Coleman et al., 1998; Millis et al., 1995; Millis et al., 2007). Consistent with previous research, the equations generated in this study also showed considerable merit for detecting malingering and correctly classifying patients with head injuries. The full multivariate model using the CVLT-II yes/no recognition errors (i.e., Pr1) correctly classified 101 out of 113 cases (88.5%) with a cut-off value of .38 on the moderate-severe TBI sample, yielding 64.5% sensitivity and 98.8% specificity. Using a more stringent cut-off, the model correctly classified 87.6% of the cases and was associated with no false positive errors (i.e., 100% specificity). In terms of the complicated mild samples, a cut-off score of .41 correctly classified 84.2% of cases putting forth good effort and 82.6% of cases having failed at least two effort measures. Using a cut-off of .69 increased the specificity for the effortful group but not the group with cases who failed two effort measures. For that group, the cut-off score had to be increased to .96 in order to achieve specificity scores above 90%. The full model using the composite scores (i.e., Pr2) resulted in the correct classification of 90 to 93 out of 113 cases (79.6-82.3%) when using the moderate-severe TBI sample, with sensitivity varying between 25.8% and 51.6% and specificity varying between 93.9% and 100% depending on the cut-off value used. Using the complicated mild samples, specificity values ranged between 89.5% and 100% for cases putting forth adequate effort and between 82.6% and 95.7% for cases who failed two or more effort measures.

Overall, both multivariate equations yielded comparable classification accuracies for all samples, except for the moderate-severe TBI sample in which Pr1 was better at differentiating between the reference and criterion groups. Given that the composites were made up of the foils, it would be expected that the two equations would yield identical classification rates. The differences in classification rates between the equations may be due to the effects of some univariate predictors cancelling each other out when the individual predictors (i.e., foils) are pooled into composites; that is, the pattern of responses that was captured in the first model was partially lost in the second model. More specifically, examining the Wald statistics from the first model (i.e., Table 8) makes it evident that UN is the only predictor whose endorsement is associated with a higher likelihood that the participant is putting forth suboptimal effort. Conversely, the other three foils have the opposite effect—that is, endorsing these foils is associated with a lower likelihood that the person is supplying suboptimal effort. Thus, by pooling the BN and UN predictors into the ETR composite, some of the predictive power from each individual predictor is lost, resulting in lower classification rates across cut-off scores. Even with this purported loss in predictive power, however, the PPP values of the second equation ranged from 76.1% to 100% depending the cut-off used, suggesting that this equation was still good at predicting malingering status. As shown in Table 23, both equations were also adequate at differentiating between malingerers and patients with complicated mild TBI putting forth adequate effort and, to a lesser degree, patients with complicated mild TBI having failed at least two effort measures.

As mentioned above, PPP values indicate the probability that someone is exaggerating or feigning symptoms given a positive test result (i.e., scoring at or above selected cut-offs). Conversely, NPP values denote the probability that an individual is supplying adequate effort given scores below selected cut-offs. Both of these values are affected by the base rates of the condition as well as by the diagnostic accuracy of the tool used to detect the condition. In general, PPP increases with higher base rates and more stringent diagnostic cut-offs whereas the opposite pattern is observed for NPP values. The base rates (i.e., prior probabilities) varied throughout the study depending on the analyses conducted and the samples used. The original sample had a base rate of malingering of 27.4%, which is close to the values reported by Mittenberg et al. (2002) for personal injury or disability claims. Given this base rate, all of the predictors displayed adequate PPP and NPP values, except for the BN predictor using a cut-off score of 4, which resulted in a PPP value of 50.1%, which hovers around chance levels and yields limited clinical utility. Conversely, the UN variable and Pr1 equation yielded the best overall predictive power values with PPPs ranging from 71.3% to 100% suggesting that there is moderate to very high probability of suboptimal effort output given an individual’s endorsement of these items. Likewise, these two variables yielded NPP values ranging from 84.5% to 88% depending on the cut-off scores used, indicating high probabilities of identifying genuine moderate-severe TBI.

In terms of the complicated mild TBI samples, the PPP and NPP values were not as high as those observed with the moderate-severe TBI sample. These results may have been affected by a combination of higher base rates (i.e., 62% for the MTBI group and 57.4% for the SE group), lower overall classification accuracy, and/or lower sample sizes. In addition, there was a discrepancy in classification accuracy between the complicated mild TBI samples, whereby the MTBI group resulted in higher PPP and NPP values than the SE group. This discrepancy may be due to a shortcoming in the predictors' utility to detect patients with complicated mild TBI giving poor effort or may be due to a sampling bias given that this group was older than the other groups and age has been previously reported to have a negative effect on CVLT-II scores (Delis et al., 2000). In order to elucidate which of these or any other factors may be responsible for the observed discrepancy, further research involving malingering and TBI should focus on comparing many different patient and litigant samples with combinations of varying degrees of head injury, types of secondary gain, degree of effort output, and base rates.

Following the procedures recommended by Greve and Bianchini (2004), hypothetical base rates were artificially generated in order to determine the utility of predictors and their respective cut-offs across different settings with varying prevalence rates. As shown in Table 13, Pr1 displayed the best predictive power across base rates and cut-off scores. This equation had very high PPP and NPP values, suggesting that participants scoring above selected cut-offs have a very high probability that they are putting forth suboptimal effort, whereas those scoring below these cut-offs have a high probability that they are supplying good effort. Similarly, the UN variable yielded PPP values of 100% and NPP values ranging from 67.4% to 95.8% across base rates when using a cut-off score of 2 or higher. The remainder of predictors had PPP values below 50% for the lowest cut-off score for a base rate of 10%, but these PPP values improved as diagnostic cut-offs became more stringent and base rates increased. The BN variable had limited utility at the lowest cut-off scores and at base rates of 10% through 30%. Overall, the best predictive values were obtained at intermediate cut-offs for base rates ranging from 20% to 40%. In the end, each predictor varies in utility depending on the degree of stringency required by the clinician as well as the setting and the purpose of testing. For example, in clinical settings with base rates around 10%, a clinician may be interested in ruling out suboptimal effort output by selecting a lower cut-off score in order to achieve higher NPP values, whereas an examiner in a medico-legal setting with base rates approaching 30% may focus on cut-off scores that yield higher PPP values in order to increase the probability of detecting malingering.

No predictor is perfect. As mentioned above, the predictors used in this study tended to misclassify moderate-severe TBI patients as putting forth suboptimal effort when these patients committed many false positive errors, whereas these predictors tended to misclassify malingerers as having moderate-severe head injuries when these participants endorsed few false positive errors. Following this finding, post-hoc ratio analyses were conducted in order to determine whether there were identifiable patterns of responses within the yes/no recognition trial that could better distinguish between patients with moderate-severe TBI and litigants putting forth suboptimal effort. Three ratios were generated from the pattern of Beta weights obtained from the multivariate analyses whereby positive Beta weights made up the numerator and negative Beta weights made up the denominator—namely, UN/PR, UN/(PR+BN+BS), and ETR/DTR. Of these, the UN/PR ratio resulted in the highest overall classification rates, ranging from 90.7% to 95.4% and surpassing all other multivariate or univariate predictors. The UN/PR cut-off score that best distinguished between the MND and moderate-severe TBI groups was .30 (i.e., a ratio of 3 UN to 10 PR). This cut-off score correctly classified 82 out of 86 cases (95.4%) and resulted in 87% sensitivity and 98.4% specificity. The UN/(PR+BN+BS) and ETR/DTR ratios also showed diagnostic promise, although their classification rates were not as good as those observed with the UN/PR ratio.

Of all the predictors analyzed in this study, both multivariate equations, the UN variable, the ETR composite, and all three ratios showed merit in distinguishing between litigants putting forth suboptimal effort and patients with genuine head injuries, although some adjustments in cut-off values were required to correctly classify the complicated mild TBI cases who had failed effort measures. The best predictors and their respective cut-off scores appear to be the UN/PR ratio at .30, Equation 1 at .40, and the UN variable at 2. Participants scoring at or above these cut-off scores had a high probability that they were putting forth suboptimal effort, whereas those scoring below these values were likely to have genuine head injuries.

Which index should examiners use? Although all these variables yielded very high predictive values, each of them may be used under different situations. Specifically, the UN variable is simple to calculate and allows the clinician to make preliminary diagnostic decisions quickly in order to maximize the utility of the assessment. Although the items making up the UN variable are dispersed throughout the yes/no recognition trial, this predictor may nonetheless be prone to coaching due to its easily identifiable item content; that is, a litigant may be instructed to be vigilant of novel and semantically unrelated items and to avoid endorsing those items when they arise. Response coaching may be encountered whenever easily tabulated cut-off scores are published (Ben-Porath, 1994). A way to combat susceptibility to coaching of a given diagnostic predictor is to generate multivariate equations that are too complex to calculate mentally while taking the test. In addition to their effectiveness in preventing coaching, multivariate equations tend to have a wider range of possible values and a more normal distribution than univariate predictors, which allow the clinician to determine finer cut-off scores to evaluate suboptimal effort with varying degrees of certainty depending on the needs of their setting. By applying various different cut-off scores, the clinician can grade an examinee's performance on a continuum according to different probabilities of negative response bias rather than according to an overly simplistic *effortful-not effortful* dichotomy, which obviates the possibility to classify degrees of mixed performance. Furthermore, because these equations incorporate many predictors at the same time, they have the potential of accounting for more variance in the dependent variable than univariate approaches. In fact, one of the multivariate equations generated in this study (i.e., Equation 1) resulted in better classification accuracy than any univariate option, including the UN variable. However, the multivariate equations' utility in combating coaching is also their main drawback. Due to their complexity, they are difficult to calculate as the assessment is progressing, which makes assessing negative response bias on the fly less feasible, thus hindering the possibility for a clinician to substitute and add tests to the assessment battery as they see fit during the evaluation.

Ultimately, the best predictor generated in this study seems to be the UN/PR ratio. It had better accuracy than the multivariate equations and univariate predictors, and it is simple enough for the clinician to calculate during testing, while remaining complex enough for the examinee to have difficulty monitoring their responses to generate ratios that fall below cut-off scores. Unfortunately, ratios cannot be used in all situations. The caveat to using ratios is that they cannot have a denominator of zero, which reduces the opportunities for their use. In the event of encountering such a scenario, reverting to the raw scores of individual foils, composites, or their respective multivariate equations would be indicated. Ideally, additional indices built into the CVLT-II that do not rely on recognition foils can be used to supplement these predictors. For example, a good practice when examining responses to the CVLT-II yes/no recognition trial would be to evaluate the number of UN or ETR foils endorsed in conjunction with the number of correct responses (i.e. hits). Endorsing a high number of UN (i.e., > 1) or ETR foils (i.e., > 5) in light of few hits (i.e., < 11) is a very unlikely pattern of responses in patients giving good effort and should alert the clinician to be suspicious of the test results, especially when there is a potential secondary gain present. Alternatively, the clinician may wait until a natural break in testing to calculate one of the aforementioned ratios, or wait until a longer break to calculate Equation 1 or 2, and then use these variables in conjunction with the number of hits to determine effort output. Using any of these techniques increases the chances that examinees giving poor effort will be detected and those with genuine head injuries giving good effort will be classified as such. It is recommended, however, that additional SVTs or effort measures derived from other neuropsychological tests be used to supplement indices derived from the CVLT-II. The purpose for this caveat is to reduce the chances of misclassifying an examinee’s performance, because predictors generated from one measure are likely highly correlated.

In any case, a single measure of effort is not recommended when making decisions about potential symptom exaggeration or suspicious performance. More specifically, although sensitivity on single measures tends to be low, as more effort tests are added to a battery, sensitivity increases multiplicatively (Iverson & Binder, 2000), which reduces the number of false negative errors. In addition, using several embedded indices reduces the false positive error rate, especially when the clinician uses indices with 90% specificity or higher. Using this "gold standard" to determine suboptimal effort would yield at most a 10% false positive rate or 1 out of 10 individuals who are tested. This false positive rate can be reduced tenfold by adding one more index using cut-off values at 90% specificity provided that the effort measures used are uncorrelated (Boone, 2007). In such a case, a person having failed two effort measures would have a 1% chance (i.e., 1/10 x 1/10) of having been misclassified as putting forth suboptimal effort when they were in fact delivering good effort. For every additional measure used, the false positive rate would decrease tenfold (i.e., 1/10). Overall, it is always recommended to use at least two indices or stand-alone tests to assess effort. Because most effort measures are at least partially correlated, the safest practice would be to use at least three effort measures to detect suboptimal effort output, as this practice has previously resulted in very low false positive errors (Larrabee, 2008).

Moreover, because some malingerers may choose to perform poorly on cognitive domains apart from memory, it is recommended that several effort indices derived from various domains be used in order to better identify sophisticated malingering practices. For example, someone may be feigning visual impairment in order to avoid military service, whereas another person may be exaggerating motor impairment to attain worker's compensation, and another person may be feigning cognitive impairment in order to receive academic accommodations. Although some of these cases might also show up as exaggerated memory impairment, only specialized indices will be able to detect malingering in these domains. In the end, however, SVTs and embedded effort indices should be used as additional tools to assist the clinician in assessment and should never replace clinical judgment.

By definition, malingering is the volitional act of symptom exaggeration or negative response bias for the purpose of achieving an identifiable secondary gain, which in this study took the form of personal injury or worker's compensation litigation status. This study was not intended to generate an embedded index to detect malingering. The embedded indices generated herein are measures of negative response bias and decisions made from their use should focus on classifying degrees of effort output, not malingering status. Although the distinction between malingering and negative response bias can sometimes seem difficult to establish, the presence of a secondary gain and volition are what differentiate the former from other conditions that involve symptom exaggeration (e.g., somatoform disorders) or low effort output and motivation (e.g., depression). When determining whether a participant is exaggerating symptomatology, it is also important to consider that other factors either in isolation or in combination can negatively influence performance on neuropsychological tests. These include but are not limited to sleep disturbances, substance abuse, psychological trauma, personality style, drug effects, and other situational factors (e.g., pressure from a partner or caregiver to manifest as more impaired in order to receive financial assistance). Although the latter example may still be considered volitional in nature, depending on the person's family dynamic, cultural background, personality, and life experiences, the individual may feel as if they have no choice but to comply with those pressures. A good clinician would have information regarding most of these factors through a thorough clinical interview, collateral interview (hopefully more than one), and chart review. Discrepancies between performance and clinical observations, report on interview, and medical records should always be evaluated in conjunction with test data to determine the presence negative response bias and, in turn, malingering. In most cases, a good clinician will consider all factors available before deciding if the person is malingering or if the performance is the result of any of the aforementioned factors. Ultimately, the label of malingering is not as important as the integrity of the neuropsychological results. All that matters is whether or not the data acquired via neuropsychological testing are valid and can be interpreted with confidence.

Finally, the purpose of this study was not to cast judgment on participants' behaviour but rather to cast doubt on the validity of neuropsychological profiles in the presence of non-credible performance. It does not matter why a person feigned cognitive dysfunction. In most settings, it does not matter *why* the data are invalid so much as *that* they are invalid. The end result is the same: the neuropsychological profile becomes uninterpretable and decisions made on such profiles must be made with caution or not at all, especially when the outcome of said decisions can have great ramifications on the person's treatment planning, litigation status, or discharge outcome. When in doubt, it is advisable to report the results as questionable rather than label someone's performance as malingering, which could influence future opportunities for treatment and future potential compensation/litigation following a genuine TBI.

References

Alexander, M. P. (1995). Mild traumatic brain injury: Pathophysiology, natural history, and clinical management. *Neurology, 45*, 1253-1260.

Altman, D. G. & Bland, J. M. (1994). Diagnostic tests 2: Predictive values. *British Medical Journal, 309*, 102-102.

American Congress of Rehabilitation Medicine (1993). Definition of mild traumatic brain injury. *Journal of Head Trauma Rehabilitation, 8*, 86-87.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text Revision). Washington, DC: American Psychiatric Association.

Ashendorf, L., O'Bryant, S. E., & McCaffrey., R. J. (2003). Specificity of malingering detection strategies in older adults using the CVLT and WCST. *The Clinical Neuropsychologist, 17*, 255-262.

Babin, P. R. (2002, April-June). Traumatic brain injury when symptoms don't add up: Conversion and malingering in the rehabilitation setting. *Journal of Rehabilitation, 68*(2). Retrieved from http://findarticles.com/p/articles/mi_m0825/is_2_68/ai_87078140/print

Backhaus, S. L., Fichtenberg, N. L., & Hanks, R. A. (2004). Detection of sub-optimal performance using a floor effect strategy in patients with traumatic brain injury. *The Clinical Neuropsychologist, 18*, 591-603.

Baldo, J. V., Delis, D., Kramer, J., & Shimamura, A. P. (2002). Memory performance on the California Verbal Learning Test-II: Findings from patients with focal frontal lesions. *Journal of the International Neuropsychological Society, 8*, 539-546.

Bauer, L., Yantz, C. L., Ryan, L. M., Warden, D. L., & McCaffrey, R. J. (2005). An examination of the California Verbal Learning Test II to detect incomplete effort in a traumatic brain-injury sample. *Applied Neuropsychology, 12,* 202-207.

Ben-Porath, Y. S. (1994). The ethical dilemma of coached malingering research. *Psychological Assessment, 6,* 14-15.

Benton, A. L., Hamsher, K. deS., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment: A clinical manual.* New York: Oxford Press.

Bianchini, K. J., Curtis, K. L., & Greve, K. W. (2006). Compensation and malingering in traumatic brain injury: A dose-response relationship? *The Clinical Neuropsychologist, 20,* 831-847.

Bigler, E. D. (1988). *Diagnostic clinical neuropsychology*. Austin, TX: University of Texas Press.

Bigler, E. D. (2001). The lesion(s) in traumatic brain injury: Implications for clinical neuropsychology. *Archives of Clinical Neuropsychology*, *16*, 95-131.

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

Blumenfeld, H. (2002). *Neuroanatomy through clinical cases*. Sunderland, MA: Sinauer.

Boccaccini, M. T., Murrie, D. C., & Duncan, S. A. (2006). Screening for malingering in a criminal–forensic sample with the Personality Assessment Inventory. *Psychological Assessment*, *18*, 415-423.

Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment* (pp. 29-49). New York: Guilford Press.

Bordini, E. J., Chaknis, M. M., Ekman-Turner, R. M., & Perna, R. B. (2002). Advances and issues in the diagnostic differential of malingering versus brain injury. *NeuroRehabilitation 17*, 93-104.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Canadian Institute of Health Information (2006). *Head injuries in Canada: A decade of change (1994–1995 to 2003–2004).* Retrieved July 6th, 2007 from http://secure.cihi.ca/cihiweb/en/downloads/analysis_ntr_2006_e.pdf

Centers for Disease Control and Prevention. (1999). *Facts about concussion and brain injury.* Atlanta, GA: Centers for Disease Control and Prevention.

Coleman, R. D., Rapport, L. J., Millis, S. R., Ricker, J. H., & Farchione, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 20,* 201-210.

Curtis, K. L., Greve, K. W., Bianchini, K. J., Brennan, A. (2006). California Verbal Learning Test indicators of Malingered Neurocognitive Dysfunction: Sensitivity and specificity in traumatic brain injury. *Assessment, 13,* 46-61.

Delis, D. C., Kaplan, E., Kramer, J. H., & Ober, B. A. (2000). *California Verbal Learning Test-II*. San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test*. San Antonio, TX: The Psychological Corporation.

Demakis (1999). Serial malingering on verbal and nonverbal fluency and memory measures: An analog investigation. *Archives of Clinical Neuropsychology, 14*, 401-410.

Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology, 9,* 80-90.

Dikmen, S. S., Machamer, J. E., & Temkin, N. R. (2001). Mild head injury: Facts and artifacts. *Journal of Clinical and Experimental Neuropsychology, 23,* 729-738.

Duncan, A. (2005). The impact of cognitive and psychiatric impairment of psychotic disorders on the Test of Memory Malingering (TOMM). *Assessment, 12*, 123-129.

Etcoff, L. M., & Kampfer, K. M. (1996). Practical guidelines in the use of symptom validity and other psychological tests to measure malingering and symptom exaggeration in traumatic brain injury cases. *Neuropsychology Review, 6*, 171-201.

Flaro, L., Green, P., & Robertson, E. (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*, 373-383.

Franzen, M. D. & Iverson, G. L. (1998). Detecting negative response bias and diagnosing malingering: The dissimulation exam. In P. J. Snyder & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment* (pp. 76-101). Washington, DC: American Psychological Association.

Frederick, R. I. & Speed, F. M. (2007). On the interpretation of below-chance responding in force-choice tests. *Assessment, 14*, 3-11.

Gennarelli, T. A. (1986). Mechanisms and pathophysiology of cerebral concussion. *Journal of Head Trauma Rehabilitation*, *1*, 23-29.

Gentilini, M., Nichelli, P., Schoenhuber, R., Bortolotti, P., Tonelli, L., Falasca, A., & Merli, G. (1985). Neuropsychological evaluation of mild head injury. *Journal of Neurology, Neurosurgery and Psychiatry, 48,* 137-140.

Goldstein, F. C., Levin, H. S. (2001). Cognitive outcome after mild and moderate traumatic brain injury in older adults. *Journal of Clinical and Experimental Neuropsychology*, *23,* 739-753.

Goleburn, C. R. & Golden, C. J. (2001). Traumatic brain injury outcome in older adults: A critical review of the literature. *Journal of Clinical Geropsychology, 7*, 161-187.

Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer administered forms, US version 1.1*. Durham, NC: CogniSyst.

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

Greenberg, D. A., Aminoff, M. J., & Simon, R. P. (2002). *Clinical neurology* (5th ed.). New York: McGraw-Hill.

Greiffenstein, M. F. & Baker, W. J. (2006, February). *Validity failure rates in a large Worker's Compensation sample with head-neck injury*. Poster presented at the Annual Meeting of the International Neuropsychological Society, Boston, Massachusetts.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*, 218-224.

Greve K. W., & Bianchini, K. J. (2002). Using the Wisconsin Card Sorting Test to detect malingering: An analysis of the specificity of two methods in nonmalingering normal and patient samples. *Journal of Clinical and Experimental Neuropsychology, 24*, 48-54.

Greve, K. W., & Bianchini, K. J. (2004). Setting empirical cutoffs on psychometric indicators of negative response bias: A methodological commentary with recommendations. *Archives of Clinical Neuropsychology, 19*, 533-541.

Greve, K. W., Bianchini, K. J., Mathias, C. W., Houston, R. J., & Crouch, J. A. (2002). Detecting malingered performance with the Wisconsin Card Sorting Test: A preliminary investigation in traumatic brain injury. *The Clinical Neuropsychologist*, *16*, 179-191.

Grubb, R. L. & Coxe, W. S. (1974). Central nervous system trauma: Cranial. In S. G. Eliasson, A. L. Prensky, and W. B. Hardin Jr. (Eds.). *Neurological pathophysiology* (pp. 292-309). New York: Oxford University Press.

Gunstad, J. & Suhr, J. A. (2001). Efficacy of the full and abbreviated forms of the Portland Digit Recognition Test: Vulnerability to coaching. *Clinical Neuropsychologist, 15*, 397–404.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual. Revised and expanded.* Odessa, FL: Psychological Assessment Resources.

Heinly, M. T., Greve, K. W., Bianchini, K. J., Love, J., M., & Brennan, A. (2005). WAIS Digit Span-based indicators of Malingered Neurocognitive Dysfunction: Classification accuracy in traumatic brain injury. *Assessment, 12,* 429-444.

Heinly, M. T., Greve, K. W., Love, J. M., & Bianchini, K. J. (2006, February). *Sensitivity and specificity of Wisconsin Card Sorting Test indicators of potential Malingered Neurocognitive Dysfunction in traumatic brain injury.* Poster presented at the Annual Meeting of the International Neuropsychological Society, Boston, Massachusetts.

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

Iverson, G. L., & Franzen, M. D. (1994). The Recognition Memory Test, Digit Span, and Knox Cube Test as markers of malingered memory impairment. *Assessment, 1,* 323-334.

Jagger, J., Fife, D., Vernberg, K., & Jane, J. A. (1984). Effect of alcohol intoxication on the diagnosis and apparent severity of brain injury. *Neurosurgery, 15,* 303-306.

Jelicic, M., Hessels, A., & Merckelbach, H. (2005). Detection of feigned psychosis with the Structured Inventory of Malingered Symptomatology (SIMS): A study of coached and uncoached simulators. *Journal of Psychopathology and Behavioral Assessment, 28,* 19-22.

King, J. H., Sweet, J. J., Sherer, M., Curtiss, G., & Vanderploeg, R. D. (2002). Validity indicators within the Wisconsin Card Sorting Test: Application of new and previously researched multivariate procedures in multiple traumatic brain injury samples. *The Clinical Neuropsychologist*, *16*, 506-523.

Larrabee, G. J. (2003a). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*, 410-425.

Larrabee, G. J. (2003b). Detection of symptom exaggeration with the MMPI-2 in litigants with Malingered Neurocognitive Dysfunction. *The Clinical Neuropsychologist, 17*, 54-68.

Larrabee, G. J. (2003c). Exaggerated MMPI-2 symptom report in personal injury litigants with malingered neurocognitive deficit. *Archives of Clinical Neuropsychology, 18*, 673-686.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*, 666-679.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A Fake Bad Scale on the MMPI-2 for personal injury claimant. *Psychological Reports, 68*, 203-210.

Lewis, J. L., Simcox, A. M., & Berry, D. T. R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI–2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment*, *14*, 170-176.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lynch, W. J. (2004). Determination of effort level, exaggeration, and malingering in neurocognitive assessment. *Journal of Head Trauma Rehabilitation, 19*, 277-283.

Mathias, G. W., Greve, K. W., Bianchini, K. B., Houston, R. J., & Crouch, J. A. (2002). Detecting Malingered Neurocognitive Dysfunction using the Reliable Digit Span in traumatic brain injury. *Assessment, 9,* 301-308.

Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, *18*, 145-152.

Meyers, J. E., Millis, S. R., & Volkert, K. (2002). A validity index for the MMPI-2. *Archives of Clinical Neuropsychology*, *17*, 157-169.

Meyers, J. E. & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology, 18,* 261-276.

Millis, S. R. (1992). The Recognition Memory Test in the detection of malingered and exaggerated memory deficits. *The Clinical Neuropsychologist, 6*, 406-414.

Millis, S. R. (2003). Statistical practices: The seven deadly sins. *Clinical Neuropsychology, 9,* 221-233.

Millis, S. R., Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment*, *7*, 463-471.

Millis, S. R., Ross, S. R., & Ricker, J. H. (1998). Detection of incomplete effort on the Wechsler Adult Intelligence Scale –Revised: A cross-validation. *Journal of Clinical and Experimental Neuropsychology, 20,* 167-173.

Millis, S. R., & Volinsky, C. T. (2001). Assessment of response bias in mild head injury: Beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology*, *23,* 809-828.

Millis, S., Wolfe, P., Larrabee, G., Hanks, R., Sweet, J., & Fichtenberg, N. (2007, February). *The California Verbal Learning Tests—II in the detection of incomplete effort.* Poster presented at the Annual Meeting of the International Neuropsychological Society, Portland, Oregon.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094-1102.

Mittenberg, W., Theroux-Fichera, S., Zielinski, R. E., & Heilbronner, R. L. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale—Revised. *Professional Psychology: Research and Practice*, *26,* 491-498.

Moore, B.A. & Donders, J. (2004). Predictors of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury*, *18*, 975-984.

Morey, L. C. (1991). *Personality Assessment Inventory.* Odessa, FL: Psychological Assessment Resources.

Nies, K. J., & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology*, *9,* 501-552.

Pankratz, L. (1988). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 169-192). New York: The Guilford Press.

Ponsford, J., Willmot, C., Rothwell, A., Cameron, P., Kelly, A, Nelms, R., et al. (2000). Factors influencing outcome following mild traumatic brain injury in adults. *Journal of the International Neuropsychological Society, 6*, 568-570.

Rao, V. & Lyketsos, C. (2000). Neuropsychiatric sequelae of traumatic brain injury. *Psychosomatics, 41*, 95-103.

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of memory malingering (TOMM). *Psychological Assessment, 10*, 10-20.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan Neuropsychological Test Battery: Theory and clinical interpretation*. Tuscon, AZ: Neuropsychology Press.

Reitan, R. M. & Wolfson, D. (2000). *Mild head injury: Intellectual, cognitive, and emotional consequences.* Tucson, AZ: Neuropsychology Press.

Rohling, M. L., Meyers, J. E., & Millis, S. R. (2003). Neuropsychological impairment following traumatic brain injury: A dose-response analysis. *The Clinical Neuropsychologist, 17*, 289–302.

Root, J. C., Robbins, R. N., Chang, L., & Van Gorp, W. G. (2006). Detection of inadequate effort on the California Verbal Learning Test-Second Edition: Forced choice recognition and critical item analysis. *Journal of the International Neuropsychological Society, 12*, 688-696.

Ross, S. R., Millis, S. R., Krukowski, R. A., Putnam, S. H., & Adams, K. M. (2004). Detecting incomplete effort on the MMPI-2: An examination of the Fake-Bad Scale in mild head injury. *Journal of Clinical and Experimental Neuropsychology, 26*, 115-124.

Ruff, R. M., & Jurica, P. (1999). In search of a unified definition for mild traumatic brain injury. *Brain Injury*, *13*, 943-952.

Schretlen, D. J., & Shapiro, A. M. (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry, 15,* 341-349.

Schwarz, L. R., Gfeller, J. D., & Oliveri, M. V. (2006). Detecting feigned impairment with the Digit Span and Vocabulary Subtests of the Wechsler Adult Intelligence Scale - Third Edition. *The Clinical Neuropsychologist, 20*, 741–753.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for Malingered Neurocognitive Dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*, 545-561.

Slick, D. J., Tan, J. E., Strauss, E. H., & Hultsch, D. F. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology*, *19*, 465-473.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25*, 183–189.

Southeastern Michigan Traumatic Brain Injury System. (2004). *What is TBI?* Retrieved May 9, 2007 from http://www.semtbis.org/semtbis/

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.). New York, Oxford University Press.

Suhr, J. A., & Boyer, D. (1999). Use of the Wisconsin Card Sort Test in the detection of malingering in student simulator and patient sample. *Journal of Clinical and Experimental Neuropsychology, 21*, 701-108.

Suhr, J. A. & Gunstad, J. (2000). The effect of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology, 15*, 415-424.

Sweet, J. J., Wolf, P., Sattlberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S., et al. (2000). Further investigation of traumatic brain injury versus insufficient effort with the California Verbal Learning Test. *Archives of Clinical Neuropsychology, 15*, 105-113.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, Allyn & Bacon.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D., F. (2002). How’d they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist, 16*, 495-505.

Tate, R L., Harris, R. D., Cameron, I. D., Myles, B. M., Winstanley, J. B., Hodgkinson, A. E., et al. (2006). Recovery of impairments after severe traumatic brain injury: Findings from a prospective, multicentre study. *Brain Impairment, 7*, 1-15.

Teasdale, G. & Jennett, B. (1974). Assessment of coma and impaired consciousness: A practical scale. *Lancet, 2*, 81-84.

Tombaugh, T. N. (1996). *Test of Memory Malingering manual.* New York: Multi-Health Systems.

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*, 260-268.

Vanderploeg, R. D., Crowell, T. A., & Curtiss, G. (1999). Verbal learning and memory deficits in traumatic brain injury: Encoding, consolidation, and retrieval. *Journal of Clinical and Experimental Neuropsychology, 23*, 185-195.

Warrington, E. K. (1984). *Recognition Memory Test*. Windsor, UK: NFER-Nelson.

Wechsler, D. A. (1981). *Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.

Wechsler, D. A. (1997). *Wechsler Adult Intelligence Scale—III*. New York: Psychological Corporation.

Whyte, J. & Rosenthal, M. (1988). Rehabilitation of the patient with head injury. In J. A. DeLisa (Ed.), *Rehabilitation medicine: Principles and practice* (pp. 585-611). Philadelphia: J. B. Lippincott Publishing.

Widows, M. R., & Smith, G. P. (2004). SIMS: Structured Inventory of Malingered Symptomatology professional manual. Odessa, FL: Psychological Assessment Resources, Inc.

Wise, J., Oliveira, M., Lacy, M., Han, D., & Pyykkonen, B. A. (2006, February). *Neuropsychological malingering of mild TBI: Effects of internet based coaching.* Poster presented at the Annual Meeting of the International Neuropsychological Society, Boston, Massachusetts.

Youngjohn, J. R., Lees-Haley, P. R., & Binder, L. M. (1999). Comment: Warning malingerers produces more sophisticated malingerers. *Archives of Clinical Neuropsychology*, *14*, 511–515.

Ziino, C., & Ponsford, J. (2005). Measurement and prediction of subjective fatigue following traumatic brain injury. *Journal of the International Neuropsychological Society*, *11*, 416-425.

Appendix

PPP can be defined as,

$$\frac{\text{True Positives}}{\text{True Positives + False Positives}} \tag{A1}$$

or, stated differently,

$$\frac{\text{(Sensitivity)(Prevalence)}}{\text{(Sensitivity)(Prevalence) + (1-Specificity)(1-Prevalence)}} \tag{A2}$$

NPP can be defined as,

$$\frac{\text{True Negatives}}{\text{True Negatives + False Negatives}} \tag{A3}$$

or, alternatively,

$$\frac{\text{(Specificity)(1-Prevalence)}}{\text{(Specificity)(1-Prevalence) + (1- Sensitivity)(Prevalence)}} \tag{A4}$$

Vita Auctoris

NAME:	Matias Ezequiel Mariani
PLACE OF BIRTH:	Rosario, Santa Fe, Argentina
YEAR OF BIRTH:	1978
EDUCATION:	Brebeuf College High School, North York, Ontario
	1992-1997
	York University, North York, Ontario
	1997-2002 B.Sc. (Specialized Honours, Psychology)
	University of Windsor, Windsor, Ontario
	2003-2005 M.A. (Clinical Neuropsychology)
	University of Windsor, Windsor, Ontario
	2005-2009 Ph.D. (Clinical Neuropsychology)