

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

The Detection of Biased Responding on the Wechsler Memory Scale- III and
Wechsler Adult Intelligence Scale- III

By

Magali Marie-Pierre Brulot,
B.A., McGill University, 1991
M.A., University of Victoria, 1996

A Dissertation Submitted in Partial Fulfilment
of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Psychology

We accept this dissertation as conforming to the required standard:

Esther Strauss, Ph.D., Supervisor, (Department of Psychology)

Catherine A. Mateer, Ph.D., Department Member, (Department of Psychology)

Helena Kadlec, Ph.D., Department Member (Department of Psychology)

Max Uhlemann, Ph.D., Outside Member (Department of Education Psychology
and Leadership Studies)

Christopher Paniak, Ph.D., External Examiner (Glenrose Hospital, Edmonton, Alberta)

© Magali Marie-Pierre Brulot, 2001
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or part, by
photocopying or other means, without the permission of the author.

Abstract

Growing demand on the limited resources available to head-injured individuals, emphasizes the need for accurate diagnosis and proper allocation of funds. Consequently, neuropsychologists are increasingly asked to render opinions regarding the validity of cognitive deficits reported following head injury. Detection of biased responding has typically been approached through the use of symptom validity measures and/ or evaluation of performance patterns on standardized neuropsychological tests.

This dissertation examined patterns of malingered performance on the Wechsler Adult Intelligence Scale-III (WAIS-III), Wechsler Memory Scale-III (WMS-III), and a self-report measure of physical and psychological symptoms. In addition, attempts were made to address several methodological concerns noted in previous analogue studies (e.g., allocation of preparation time). Malingered performance was compared to that of a normal control group (NC = 34) and a group of mildly head injured individuals (MHI = 22). Results revealed that the simulating group (SIM = 32) endorsed more subjective concerns than the NC group. On the cognitive measures, simulators showed a tendency towards general suppression of performance versus specific areas of deficit (e.g., attention). Specifically, the SIM group suppressed their performance on the WAIS-III, but not typically enough to differentiate them statistically from either the NC or MHI groups. The SIM group's performance on the WMS-III was more in keeping with the overall suppressed performance pattern reported in previous research. Although simulators often performed significantly worse than the NC group, they did not generally suppress their performance excessively when compared to the MHI group. Results

obtained in this study may reflect the use of more detailed instructions or the preparation time allocated to the SIM group. Overall, these findings emphasize the importance of incorporating multiple detection measures to assure the accurate assessment and diagnosis of head injury.

Examiners:

Esther Strauss, Ph.D., Supervisor, (Department of Psychology)

Catherine A. Mateer, Ph.D., Department Member (Department of Psychology)

Helena Kadlec, Ph.D., Department Member (Department of Psychology)

Max Uhlemann, Ph.D., Outside Member (Department of Education Psychology and Leadership Studies)

Christopher Paniak, Ph.D., External Examiner (Glenrose Hospital, Edmonton Alberta)

Table of Contents

Abstract.....	ii
Table of Contents.....	iv
List of Tables.....	vii
List of Figures.....	viii
Acknowledgements.....	ix
Dedication.....	x
Introduction.....	1
Role of Neuropsychologists.....	1
Head Trauma.....	2
Malingering.....	4
Prevalence of Biased Responding.....	8
Approaches to the Study of Malingering.....	9
Incentives.....	11
Knowledge.....	13
Assessment of Biased Responding.....	15
Performance Patterns of Simulators.....	20
Proposed Study.....	26
Methodological Considerations.....	26
Hypothesis I: Subjective Complaints.....	27
Hypothesis II: Performance on WAIS-III and WMS-III.....	27
Hypothesis III: Comparison of WAIS-III and K-BIT Performance.....	29

Hypothesis IV: Consistency of Performance.....	30
Hypothesis V: 21-Item Test.....	30
Methods.....	31
Participants.....	31
Design and Procedure.....	34
Control Group Instructions.....	34
Simulating Group Instructions.....	34
Materials.....	36
BCPSI.....	36
21 Item Test.....	37
K-BIT.....	37
WAIS-III.....	38
WMS-III.....	38
Results.....	40
Part I: Subjective Complaints.....	41
Part II: WAIS-III Summary Scores.....	44
Part III: WAIS-III Discrepancy Scores.....	47
Part IV: WAIS-III Subtests.....	48
Part V: Comparison of WAIS-III and K-BIT.....	53
Part VI: WMS-III Summary Scores.....	54
Part VII: WMS-III GMI-WMI Discrepancy Score.....	57
Part VIII: WMS-III Subtest Scores.....	58
Part IX: Reliable Digit Span and Reliable Spatial Span.....	62

Part X: Consistency of Performance.....	63
Part XI: 21-Item Test.....	63
Discussion.....	65
Conclusion.....	69
References.....	74
Footnotes.....	84

List of Tables

Table 1. Differential Diagnosis of Malingering and Factitious Disorder.....	7
Table 2. Problems with Simulation Studies Using the Wechsler Scales.....	23
Table 3. Demographic Information on NC, SIM, and Litigating Groups.....	33
Table 4. BCPSI Summary Scores.....	42
Table 5. BCPSI Psychological and Physical Symptoms Reported by the NC and SIM Groups.....	43
Table 6. Means and Standard Deviations for WAIS-III Summary Scores and Discrepancy Scores.....	45
Table 7. Means and Standard Deviations for WAIS-III Age-Adjusted Subtest Scores.....	49
Table 8. Mean and Standard Deviations for WAIS-III Subtest Raw Scores Controlling for Age and Education.....	52
Table 9. Means and Standard Deviations for WAIS-III & K-BIT Summary Scores.....	54
Table 10. Means and Standard Deviations for WMS-III Summary Scores.....	56
Table 11. Means and Standard Deviations for WMS-III Age-Adjusted Subtest Scores.....	59
Table 12. Mean and Standard Deviations for WMS-III Subtest Raw Scores Controlling for Age and Education.....	61
Table 13. Means and Standard Deviations for Reliable Digit Span and Reliable Spatial Span.....	62
Table 14. Means and Standard Deviations for 21-Item Test Scores.....	64

List of Figures

Figure 1. Group Profiles on WAIS-III Summary Scores.....	46
Figure 2. Group Profiles on WMS-III Summary Scores.....	55
Figure 3. Group Profiles on WMS-III Subtests.....	60

Acknowledgements

The author would like to thank Dr. Esther Strauss, Dr. Helena Kadlec, Dr. Grant Iverson, and Dr. Michael King for their valuable contributions to this project. In addition, a special thanks is extended to Stuart MacDonald for his time and patience.

Dedication

This dissertation is dedicated to my parents and family for their never ending support, to the Sawchyn family for their encouragement, and especially to Jim Sawchyn for his love, patience, and support.

The Detection of Malingered Closed-Head-Injury on the Wechsler Memory Scale-III and
Wechsler Adult Intelligence Scale-III

In the past decade, there has been a growing interest in the ability to detect biased responding on neuropsychological measures (Bernard, McGrath, & Houston, 1993). Improved vehicular safety measures have resulted in increased survival rates following motor vehicle accidents, and consequently, more demand on the limited resources available to head injured individuals. With ongoing changes in health care management, long-term treatment of patients has become increasingly rare (Haines & Norris, 1995). The importance of proper diagnosis and the allocation of the limited available resources to individuals who truly need them, is therefore paramount (Franzen, Iverson, & McCracken, 1990). In addition, the cost of malingering to society is deceptively large (Price & Stevens, 1997). A recent investigation by the Coalition Against Insurance Fraud estimated the total cost of insurance fraud in the United States at a staggering \$85.3 billion in 1995 (LoPiccolo, Goodkin, & Baldewicz, 1999). It is thus not surprising that the discrimination of individuals with genuine head injury sequelae from those exaggerating or faking deficits, is becoming an integral part of neuropsychological assessments.

Role of Neuropsychologists

Neuropsychologists are increasingly being called upon to determine the extent of cognitive deficits following head injury and to make suggestions regarding treatment planning (Gutierrez & Gur, 1998). In addition, in the medical-legal context, neuropsychologists are expected to render opinions regarding the validity of patients' cognitive symptoms (Nies & Sweet, 1994). When clinicians fail to detect biased

responding, their professional credibility decreases (Haines & Norris, 1995). The methodology by which conclusions are reached is often challenged in the medical-legal setting, thereby emphasizing the need to establish empirical methods for determining the validity of test results (Mittenberg, Theroux-Fichera, Zielinski, & Heilbronner, 1995).

Neuropsychologists typically use a combination of clinical interviews, file review, corroborative information, and a variety of standardized neuropsychological tests, to evaluate an individual's cognitive and emotional functioning following head injury. With regards to standardized tests, a comprehensive neuropsychological assessment usually includes the administration of a test of intelligence and memory measures (e.g., Wechsler Adult Intelligence Scale-III (WAIS-III; Psychological Corp., 1997a) and Wechsler Memory Scale-III (WMS-3; Psychological Corp., 1997b)). The earlier edition of the Wechsler intelligence test (WAIS-R) has been reported to be used in over 90% of neuropsychological evaluations (Guilmette, Faust, Hart, & Arkes, 1990; Mittenberg et al., 1995), and has been found to be sensitive to the cognitive effects of cerebral trauma (Crossen & Wiens, 1988; Rawlings & Crewe, 1992; Wechsler, 1997).

Head Trauma

The effects of head trauma are varied, depending upon the structures involved. While open head injury may result in focal damage, closed head injury is typically associated with more diffuse or generalized impairment (see Lezak, 1995 for review). In closed head injury (CHI) there is no penetration of the skull. The force of the trauma (e.g., motor vehicle accident, assault with blunt object) exerts an impact on the brain within the closed, bony space of the skull, often resulting in frontal and temporal injuries (Reitan & Wolfson, 1985). The resulting deficits may include organizational difficulties

(Levin & Goldstein, 1986; Zappala & Trexler, 1992), memory problems (Levin, 1989; Reitan & Wolfson, 1985), attention and concentration difficulties (Kay, Newman, Cavallo, Ezrachi, & Resnick, 1992), as well as a decline in cognitive functions affecting overall intelligence (Rawlings & Crewe, 1992; Reitan & Wolfson, 1985). In most cases, there is a combination of focal and diffuse damage such that specific deficits blend into a more general pattern of impairment (Miller, 1990). As a result, an uneven pattern of cognitive performance may emerge, with some abilities severely impaired, others only mildly affected, and still others apparently well preserved; all superimposed on general overall "sluggish" cognitive processing (Miller, 1990). These deficits may be permanent or temporary depending on the severity of the head injury.

Numerous studies have documented the relationship between severity of head injury, and the degree and persistence of deficits (see Levin, 1989). In the CHI population, the assessment of individuals who sustain mild traumatic head injury can be particularly challenging for neuropsychologists (Ruff & Richardson, 1999), as subtle and often transient symptoms typically result (Binder, Rohling, & Larrabee, 1997). The collection of symptoms often seen following mild head injury has been termed the post-concussion syndrome (PCS) and can include cognitive complaints (attention, concentration, and memory difficulties), physical complaints (headaches, fatigue, dizziness, blurred vision, sensitivity to light and noise), and psychosocial complaints (irritability, depression, anxiety, personality changes (American Psychiatric Association, 1994). These symptoms have been reported in other clinical populations (e.g., Fox, Lees-Haley, Earnest, & Dolezal-Wood, 1995) and are also found present in the general population (Gouvier, Uddo-Crane, & Brown, 1988), suggesting that other factors (e.g.,

emotional factors, adjustment problems, effects of medications) may contribute to their manifestation.

Although most individuals with a mild head injury recover within the first three months following their head trauma (Binder et al., 1997; Dikmen, McLean, & Temkin, 1986), a small number of patients report persisting symptoms that are sometimes of a severity disproportionate to the injury sustained (Suhr, Tranel, Wefel, & Barrash, 1997).

These deficits following minor head injury can occur in the context of otherwise normal neurological/neuroradiological findings (Gronwall, 1991). In such instances, the neuropsychological evaluation may provide the only evidence suggesting the presence of sequelae following traumatic head injury (Trueblood & Frank, 1994). It is at such times that the validity of the neuropsychological test results is most often questioned, and the issue of biased responding (e.g., malingering) raised.

The underlying causes of biased responding may be viewed as falling along a continuum (Haines & Norris, 1995) ranging from unconscious production of symptoms (e.g., Somatoform disorders), to intentional symptom fabrication to meet intrapsychic needs (e.g., Factitious disorder), to exaggeration of symptoms actually present (e.g., cry for help), to deliberate fabrication of symptoms for external incentives (e.g., malingering). Research on biased responding over the past 15 years, has typically focussed on malingering.

Malingering

With regard to malingering, it is not known if this form of biased responding reflects a few basic and inter-knit underlying dimensions that are relatively consistent across situations, individuals, and feigned conditions, or if it is best conceptualized as

multiple independent dimensions that may vary markedly based on the person, condition feigned, or situation (Faust & Ackley, 1998). Miller and Miller (1992) state that it is unlikely that malingering is a homogenous construct, and thus emphasize the need for measures sensitive to malingering in various cognitive domains (e.g., visuospatial ability, memory). The homogeneity of malingering or its structure, if it has one, will bear greatly on the method used for its identification (Faust, & Ackley, 1998). While the presence of malingering does not necessarily mean that injury or true dysfunction is absent, it does invalidate the test results (Trueblood & Frank, 1994).

The fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (American Psychiatric Association, 1994) does not present formal criteria for the diagnosis of malingering. Rather, it describes malingering as the intentional production of false or grossly exaggerated symptoms motivated by external incentives (e.g., financial compensation or evading prosecution), and suggests four indices that should raise suspicion to the possibility of malingering. In particular, malingering should be strongly suspected if any combination of the following is noted: (1) medical-legal context of presentation; (2) marked discrepancy between claimed disability and objective findings; (3) lack of cooperation; and (4) presence of Antisocial Personality Disorder. It has been suggested that the discrepancy between claimed deficits and difficulties observed upon evaluation, is by far the most reliable context which should lead one to consider the possibility of malingering (LoPiccolo, et al., 1999).

As Slick and colleagues (1999) point out, however, a consensus has yet to be reached on the definition and criteria for malingering. With regard to the application of this label in the neuropsychological context, Greifenstien, Baker, and Gola (1994) proposed a set of criteria for the diagnosis of overt malingering of memory deficits and

demonstrated clinically significant association between classification using these criteria and performance on tests specifically assessing for biased responding. Their proposed criteria consisted of (1) improbably poor performance on two or more neuropsychological measures, (2) total disability in a major social role, (3) contradiction between historical information provided and material derived from collateral sources, and (4) remote memory loss.

The above mentioned criteria, however, are restricted to memory deficits and do not include a definition of malingering, differential diagnoses, or behavioural observations (Slick et al., 1999). To address these concerns, Slick and colleagues (1999) proposed the following definition and criteria, combinations of which can be used to specify the certainty of the diagnosis (see Slick et al., 1999):

Definition:

Malingering of Neurocognitive Dysfunction (MND) is the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility. Substantial material gain includes money, goods, or services of non-trivial value (e.g., financial compensation for personal injury). Formal duties are actions that people are legally obligated to perform (e.g., prison, military, public service, child support, or other financial obligations). Formal responsibilities are those that involve accountability or liability in legal proceedings (e.g., competency to stand trial).

Diagnostic Criteria:

- A) Presence of substantial external incentive
- B) Evidence from neuropsychological testing
 - 1) Definite or probable response bias
 - 2) Discrepancy between test data and known patterns of brain dysfunction
 - 3) Discrepancy between test data and behaviour observed
 - 4) Discrepancy between test data and reliable collateral reports
 - 5) Discrepancy between test data and documented background history
- C) Evidence from self-report data
 - 1) Self-report history is discrepant with documented history
 - 2) Self-report symptoms are discrepant with known patterns of brain functioning
 - 3) Self-report symptoms are discrepant with observed behaviour

- 4) Self-reported symptoms are discrepant with information obtained through collateral sources
- 5) Evidence of exaggerated or fabricated psychological dysfunction
- D) Behaviours meeting necessary criteria from groups B or C are not fully accounted for by psychiatric, neurological, or developmental factors.

When considering deliberate fabrication of symptoms, the distinction between malingering and factitious disorder needs to be made. Differential considerations for these disorders (as reviewed in Overholser, 1990) are presented in Table 1.

Table 1

Differential Diagnosis of Malingering and Factitious Disorder (Overholser, 1990)

	Malingering	Factitious
Behaviour During Interview	Guarded	Guarded
Previous Hospitalizations	Occasional	Definitely
Stability of Problems	Situational	Continual
Emotional Response to Treatment	Agreeable	Belligerent
Behavioural Response	Cooperative	Uncooperative
Production of Symptoms	Conscious	Conscious
Control over Symptoms	Voluntary	Involuntary
Source of Motivation	External	Internal

While assessment techniques may be able to detect less than optimal performance, the underlying cause must be determined through careful review of the patient's history, corroborative information, context, and test results. It is also important to remember that in addition to patient related factors, inaccuracies can arise from unintentional sources

such as misadministration of tests and extraneous factors such as fatigue, medication side effects, and transient poor mood (Faust & Ackley, 1998). Interestingly, although the assessment of malingering has become prevalent, it continues to be diagnosed infrequently. Trueblood and Frank (1994) hypothesize that this may stem, in part, from the difficulty inherent in trying to maintain a balance between therapeutic alliance and healthy skepticism, concern regarding negative reactions from colleagues, and fear of reprisal (e.g., lawsuit).

Prevalence of Biased Responding

It has been suggested that the rate of biased responding is particularly high in populations where there is incentive for impaired performance (Binder, 1997). Biased responding is therefore an especially contentious issue in medical-legal settings where patients are seeking compensation for their injuries. Although the actual base rate of biased responding in the medico-legal context is unknown, some researchers have suggested that resolution of symptoms following settlement is indicative that less than optimal effort was put forth by litigants in the pre-settlement phase (e.g., Kay et al., 1992). Other investigators, however, have found that settlement of litigation is not necessarily associated with improvement of symptoms (e.g., Binder, 1986; Resnick, 1988), implying that the prevalence of biased responding (e.g., malingering) in this population is not as high as presumed. For example, Mendelson (1995) studied a group of 264 litigants who had not returned to work at the time that their case was settled. Follow-up of these individuals approximately two years later revealed that 75% of them remained unemployed.

Although research on biased responding, and in particular malingering, has greatly increased in the past decade, our knowledge in this area is still limited. Since the goal of malingering is to go undetected, the incidence of biased responding remains unknown (Nies & Sweet, 1994). Estimates of biased responding have ranged from 1% to more than 50%, based on the population studied (Greiffenstein et al., 1994; Resnick, 1988) and the astuteness and skepticism of the clinician (Binder, 1990).

Approaches to the Study of Malingering

Research on malingering has typically been approached by one of two means: (1) analysis of the performance of 'suspect' individuals (e.g., who have external incentive to put forth less than optimal effort (e.g., litigants), those who provide either an improbable symptom history, are totally disabled in at least one major life role (e.g., occupation), or who perform questionably on a measure of symptom validity (e.g., Greiffenstein et al., 1994)), and (2) research with persons simulating head injury symptoms. Although initial studies with suspected malingering often involved case reviews (e.g., Binder & Pankratz, 1987), recent studies have included larger sample sizes (e.g., Binder, 1993; Greiffenstein et al., 1994). Results from a number of these studies have suggested that performance of some patients is questionable (e.g., unusual results in light of the injury sustained) when there is financial incentive to be impaired (e.g., Binder, 1997; Binder & Rohling, 1996; Schmand et al., 1998).

The use of suspected malingering is advantageous because there are a number of features inherent to true malingering that can not be easily simulated (e.g., financial motivation, experience of a traumatic event etc.). In addition, malingering situations have negative incentives that can not be readily replicated (e.g., fear of being discovered, loss

of job, and embarrassment). However, a number of problems exist with the research done to date. Although research findings have often revealed group mean differences, many of the 'suspected' malingeringers score within the same range as brain injured patients on a variety of neuropsychological measures and malingering tests (e.g., Trueblood, 1994). In the absence of absolute evidence of malingering, it is reasonable to conclude that some of the 'suspected' malingeringers may not, in fact, be malingering (Haines & Norris, 1995; Rogers, Harrell, & Liff, 1993). Another problem with investigating 'suspected malingeringers' is that information is lacking regarding the actual base rate of malingering in certain populations. Thus, it is difficult to determine if a particular assessment method leads to increased diagnostic accuracy (Faust, Hart, Guilmette, & Arkes, 1988; Rogers et al., 1993). In addition, with the inability to identify true prevalence rates it becomes difficult to know how to interpret group differences, even when they occur in the hypothesized direction; we do not know if the group differences reflect an under-prediction or over-prediction of biased responding, or even whether the 'suspect' individuals truly represent malingeringers (Rogers et al., 1993). Due to the fundamental problems associated with studies looking at 'suspect' malingeringers in high risk populations, Rogers and colleagues (1993) suggested that this approach to the research of malingering should be discouraged.

Ideally, research should be conducted with data collected from individuals known to have malingered. Even then, however, the subjects studied would not be representative of all malingeringers since the sample would only include patients who have been caught. A representative sample should also include malingeringers who have successfully feigned a disorder or illness. However, given the nature of this problem and the inherent goal of not being detected, studies on representative samples of malingeringers are not possible.

Alternatively, research has been conducted using individuals asked to purposely fake symptoms of head injury. The primary advantage of simulation studies is the use of systematic comparisons under well-defined experimental conditions (Rogers et al., 1993). Within the experimental context, the base rate of biased responding is known. Therefore, assessment techniques that detect these subjects accurately are assured to be internally valid (Haines & Norris, 1995). However, this approach has been criticized due to the fact that variables present in genuine cases of malingering may not be easily reproduced in simulation research settings (Trueblood & Frank, 1994). For example, the motivation to malinger (e.g., large external incentive), the occurrence of a traumatic event, and contact with numerous sources of information (e.g., lawyers, care providers, other patients) are all factors that are absent in the experimental context. In addition, simulation studies to date have not taken into account other factors such as depression, anxiety and chronic pain, that may contribute to deficits reported by head injured individuals (Suhr et al., 1997). Consequently, the application of findings from simulation studies to the clinical context has been questioned (Franzen, et al., 1990; Rogers et al., 1993).

Incentives

A number of attempts have been made to improve the generalizability of findings from analogue studies, by more closely approximating real life situations of malingering. For example, a number of researchers have attempted to offer various types and amounts of incentive to experimental malingerers (e.g., Bernard, 1990; Fredrick, Sarfaty, Johnston, & Powel, 1994; Martin, Bolter, Todd, Gouvier, & Nicolls, 1993). While some researchers have found that incentive influences detection rates (e.g., Fredrick et al., 1994), others have found that regardless of the incentive offered, simulating malingerers

with motivation do not differ in performance from those not offered an incentive (e.g., Bernard, 1990; Martin et al., 1993). Although the impact of positive incentives on performance remains equivocal, studies have often found that regardless of the incentive, simulators tend to over exaggerate their deficits and consequently perform more poorly than truly HI individuals (Baker, Hanley, Jackson, Kimmance, & Slade 1993; Martin et al., 1993). However, it is not possible to determine how experimental results using smaller positive incentives, relate to the larger incentives available through litigation (Haines & Norris, 1995). Perhaps with larger positive incentives at stake, clients would approach neuropsychological evaluations more cautiously, and would thus have test results that more closely approximate those of brain-injured individuals.

In addition, few studies have explored the effects of negative incentives on test performance, although real life malingering situations are associated with numerous negative incentives (e.g., embarrassment). Research by Rogers and Cruise (1998) focussed on the use of mild negative incentives, and found that the potential loss of class credit or public posting resulted in a more concerted effort to feign specific symptoms of depression. However, research on the effects of negative incentives remains scant. Overall, incentives may not be important so much for how they alter effort in the examination setting, but rather for what they lead individuals to do in the pre-assessment phase (Faust & Ackley, 1998). When the stakes are high, an individual may spend a considerable amount of time preparing to malinger, thereby increasing their knowledge of the deficits they are attempting to feign.

Knowledge

The effect of knowledge on malingering has also been studied in an attempt to more closely approximate real-life malingering conditions. Research conducted by Aubrey, Dobbs, and Rule (1989), has indicated that lay persons have little understanding of cognitive or psychiatric symptoms commonly experienced following head injury, and therefore are unlikely to accurately simulate them (e.g., endorse highly unusual symptoms). These researchers found that 80% of undergraduates thought physical symptoms would likely result after head injury, but that less than 50% thought that head injury would result in cognitive symptoms. It has been suggested, however, that successful malingers may be sufficiently motivated to educate themselves about the disorder they are attempting to feign (Cochrane, Baker, & Meudell, 1998; Haines & Norris, 1995). In addition, as litigants typically undergo repeated evaluations, repetitive exposure to various assessment measures may help them present themselves more accurately in a cognitively compromised way (Strauss, Hultsch, Hunter, Slick, Patry, and Levy-Bencheton, 2000). Furthermore, research has also shown that some attorneys prepare litigants for an evaluation by describing the process or even telling clients how to respond (Youngjohn, Lees-Haley, & Binder, 1999). Wetter and Corrigan (1995, in Dicarlo, Gfeller, & Oliveri, 2000) found that the majority of law students and practicing attorneys in their sample stated that they would engage in coaching. In addition, Youngjohn (1995) reported on a case where the attorney admitted to coaching the client prior to the assessment.

There can be great variability in the amount and quality of information available to individuals undergoing neuropsychological evaluations. For example, clients may be warned that the assessment will include tests aimed at detecting biased responding, they

may be provided with information about common symptoms of head injury, and may even be coached on how to complete certain tasks. The client's knowledge and preparation can potentially have a significant impact of the clinician's ability to detect biased responding.

The influence of knowledge on biased responding has been studied in research using simulating malingeringers. The range of information provided has varied from minimal (Martin et al., 1993; Suhr & Gunstad, 2000) to elaborate information detailing ways to malinger effectively (see Fredrick & Foster, 1991, in Haines & Norris, 1995). The results from studies to date have been inconclusive. Some researchers have noted that the level of sophistication significantly affects test performance (DiCarlo et al., 2000; Johnson & Lesniak-Karpiak, 1997; Rose, Hall, Szalda-Petree and Bach, 1998; Suhr & Gunstad, 2000). For example, Martin and colleagues (1993) noted that 90% of sophisticated simulators scored at or above chance in comparison to only 54% of naïve malingeringers. However, although coached malingeringers may have been able to fake symptoms of head injury more accurately, they still tended to exaggerate their deficits in comparison to truly head injured individuals (Gouvier, Hayes, & Smiroldo, 1998).

Other studies, on the other hand, have failed to reveal significant differences based on the level of information provided to experimental malingeringers (e.g., Johnson, Bellah, Dodge, Kelley, & Livingston, 1998; Klimczak et al., 1997; Osimani, Alon, Berger, & Abarbanel, 1997). In addition, personal experience does not appear to significantly affect the ability to malinger effectively. A study by Hayes, Martin, and Gouvier (1995), revealed that college students with a history of mild head injury and knowledge of HI sequelae were not able to simulate more effectively. In addition, Hayward, Hall, Hunt, and Zubrick (1987) found that 28 registered nurses with varying degrees of experience

working with neurological and neurosurgical patients had great difficulty reproducing the test performance of individuals with left frontal injury. It may also be however, that the effect of coaching depends on the task an individual is attempting to feign. Alliger, Lilienfeld, and Mitchell (1996) found that coaching helped subjects on tasks with a clear purpose, but did not affect performance on tests with a disguised purpose.

While the applicability of results from simulation studies to clinical context remains to be determined, at the very least, these results can serve as a helpful starting point in evaluating malingering. Specifically, simulation studies are perhaps most useful in identifying candidate variables meriting analysis in applied clinical settings. Variables found through analogue studies have a greater probability of being applicable to clinical settings than random guess and perhaps even hypotheses based on clinical experience (Faust & Ackley, 1998). Studies looking at the generalization of MMPI indicators (Rogers et al., 1993) offer an optimistic outlook regarding the possibility that experimental variables discriminating malingeringers may generalize to the clinical setting.

Assessment of Biased Responding

Research on malingering has evaluated a number of methods used in the detection of biased responding. For example, studies have looked at the effectiveness of using clinical judgement, single tests specifically designed to assess for malingering, and identification of patterns in test performance on standardized neuropsychological tests.

Research into objective measures of biased responding has stemmed from findings that subjective techniques were often not effective; that the use of clinical experience and judgement often leads to inaccurate conclusions regarding malingering (Faust & Guilmette, 1990; Trueblood & Frank, 1994). For example, Heaton, Smith,

Lehman, and Vogt (1978) found that 10 neuropsychologists asked to blindly classify the protocols of experimental malingeringers and non-litigating head injured patients, were able to do so only at chance or slightly above chance level. In that study, however, the neuropsychologists had a wide range of experience, and were not provided with information typically used in the current evaluation of malingering (e.g., verbatim responses, behavioural observations, premorbid status, specialized tests). Nonetheless, supporting evidence for the lack of accuracy of subjective assessment of malingering also comes from the work of Fredrick and colleagues (1994). These researchers found that neuropsychologists had difficulty detecting malingering even though they had performed comprehensive evaluations involving face to face contact with the clients. In addition, Faust and colleagues (1988) found that in their sample, none of 42 neuropsychologists asked to judge malingered profiles identified biased responding. In addition, 74% of these clinicians were moderately to highly confident of their diagnosis. However, Trueblood and Binder (1997) point out that this study was conducted with adolescents and that there was therefore likely less expectancy for malingering. In their own study, they found that neuropsychologists are able to detect malingering at least in obvious cases. It should be noted that a number of the studies exploring the ability of professionals to detect biased responding were conducted over 15 years ago, in the absence of specialized tests aimed at detecting malingering .

The suggestion that clinical experience may not suffice for accurate identification of malingering is not surprising given the past 20 years of research on the accuracy of the detection of lying (e.g., Ekman & O'Sullivan, 1993). Results from studies in this area indicate that little confidence should be placed in the judgments by lay persons or experts, about whether someone is lying or telling the truth (Ekman & O'Sullivan, 1993). The

literature suggests that lay persons and professionals from various walks of life (e.g., psychologists, psychiatrist, lawyers, judges) have considerable difficulty detecting lies (De Paulo, 1994, in Franzen & Iverson, 1997), perhaps due to the susceptibility of human thinking to inconsistencies (Dawes et al., 1989 in Trueblood & Frank, 1994), or the number of sources of bias and error (Wedding & Faust, 1989). Clinicians who rely solely on experience to detect malingering would face the same conditions as someone trying to learn under conditions of sporadic, skewed, delayed, noisy, and all too often, misleading feedback (Faust & Ackley, 1998). We are prone to forming false associations between signs and disorders and often overestimate the strength of these associations.

The search for empirically validated objective measures of malingering has lead to the construction of tests specifically designed to assess biased responding. A popular approach has been to use forced choice procedures such as the Victoria Symptom Validity Test (Slick, Hopp, Strauss, Hunter, & Pinch, 1994) and 21-Item Test (Iverson, Franzen, & McCracken, 1994). In interpreting symptom validity test results, two approaches have typically been employed: one approach requires the application of the binomial theorem to test results, under the rationale that a person's performance should at least fall at chance level even if the person has no recollection of the information that was presented. Therefore, if a person's performance falls below the chance level, it is thought to reflect a deliberate attempt to do poorly on the task (Reynolds, 1998). This statistical approach has been criticized because few malingerers actually score below chance and thus it most clearly identifies the obvious malingerers (Haines & Norris, 1995; Rogers et al., 1993). As a result of this criticism, a second approach has been proposed, in which test performance is compared to cutoffs that have been established based on comparison with the poorest performance exhibited by other clinical groups (e.g., truly brain injured

individuals). For example, brain injured individuals typically pass 90% or more of recognition trials (see Bianchini, Mathias, & Greve, 2001 for a review of symptom validity measures), and thus performance falling below this level would suggest that inadequate effort may have been put forth on the task.

Overall, however, symptom validity tests have been found to have only moderate sensitivity and have consistently shown lower than optimum negative predictive power (see Bianchini et al., 2001 for review). In addition, the face validity of forced-choice procedures may also detract from their utility, as the objective to perform above chance may be realized by some clients (Haines & Norris, 1995).

Whether using forced choice procedures or other methods (e.g. Rey 15 Item Memory Test; Rey, 1964 in Spreen & Strauss, 1998) in the detection of malingering, single test approaches have also been criticized for the specificity of the material covered (Pankratz, 1983), as most of these tests assess a limited range of cognitive abilities (e.g., verbal memory). It is unlikely that actual clinical malingers would complain or fake only the abilities tapped by the single measure used (Nies & Sweet, 1994). Thus, a measure that is shown to detect malingering of verbal memory deficits, cannot be assumed to detect malingering of other problems often associated with head injury (Faust & Ackley, 1998). In addition, single test approaches do not provide information about the consistency of performance across tasks assessing similar abilities, which has been proposed as a potentially useful marker for malingering (Reitan & Wolfson, 1996). Another pitfall of the single test approach is that it may be considerably easier to fake well on one measure tapping a specific cognitive ability than on a number of tests tapping a variety of skills (Nies & Sweet, 1994). As well, with tests specifically assessing malingering, some motivational issues such as selective biased responding or certain

strategies of malingering may also not be efficiently assessed (Trueblood, 1994). As a result, some malingering individuals may not be detected by these selective procedures.

Given the cautionary statements regarding the interpretation of single test results (e.g., Lezak, 1995), it may be difficult to make sense of questionable findings on a single objective measure specifically designed to assess biased responding. Typically, clinicians do not administer only one test to assess a particular cognitive domain, as it would not be prudent to make statements regarding cognitive ability based on the results from one measure. Similarly, having obtained unremarkable results on one malingering task, the clinician should be weary about concluding that good effort was put forth throughout the test battery. This conclusion could also be misleading because the sensitivity of malingering tests is often quite poor (Rogers, Harrell & Liff, 1993; Wiggins & Brandt, 1988). Thus, it may be problematic to generalize clinical inference from the effort expended on one measure to that expended on other tests (Franzen & Iverson, 1997).

As the existence of these malingering tests becomes better known, it may become easier for a subject facing assessment to have some notion of how to feign impairment without being detected (Franzen & Iverson, 1997). Lastly, another problem with tests specifically aimed at assessing biased responding is that they require additional time and thus prolong the assessment (Suhr & Boyer, 1999). In short, while specific tests of malingering are perhaps useful when used in conjunction with other techniques, conclusions regarding motivation during the assessment based primarily on these specific tests may be difficult to defend.

Another objective approach suggested to detect malingering has been to look at performance on existing standardized neuropsychological tests (Iverson, Slick, & Franzen, 2000). This makes intuitive sense: since we are interested in the validity of

performance on these tests, why not evaluate effort throughout the testing session? One advantage to this approach is that neuropsychological tests have scoring criteria for presumed optimal effort. In addition, by using tests that are typically part of a neuropsychological evaluation, additional time for assessment of malingering is not required, thus allowing for greater efficiency in the assessment (Franzen & Iverson, 1997); if malingering is ruled out, the obtained test results can be used to evaluate brain functioning. The use of standardized tests such as the WMS-III and WAIS-III is also beneficial as these measures are well suited to the task of cataloging cognitive strengths and weakness because they sample a broad range of cognitive ability. As mentioned previously, it is most likely more difficult to successfully malinger throughout a testing session than on one test tapping a specific cognitive domain (Nies & Sweet, 1994).

Performance Patterns of Simulators

Understanding the performance patterns of malingeringers on the Wechsler scales has been of particular interesting given the frequency of the use of these measures (Guilmette et al., 1990; Mittenberg et al., 1995) in neuropsychological evaluations. Research on biased responding, and specifically malingering, has revealed a number of patterns in the performance of simulators on the WAIS-R and WMS. Specifically, simulators' overall intellectual abilities have typically been found to be suppressed, and in fact, they often appear to perform more poorly than truly brain injured individuals (e.g., Binder 1990; Binder & Willis, 1991; Mittenberg, Azrin, Millsaps, & Heilbronner, 1993; Trueblood 1994; Trueblood & Schmidt, 1993). In addition, Performance IQ has frequently been found to be lower than Verbal IQ; a discrepancy in the direction typically

observed in head injured persons (Mandleberg & Brooks, 1975; Rawlings & Crewe, 1992; Mittenberg et al., 1995).

Researchers have also found that, unlike head injured individuals, simulators tend to perform more poorly on recognition tasks than on spontaneous recall (Bernard, 1990; Binder, Villanueva, Howieson, & Moore, 1993; Brandt, Rubinsky & Lassen, 1985; Greiffenstein et al., 1994; Iverson, Frazen, & McCracken, 1991; 1985; Trueblood & Schmidt, 1993; Wiggins & Brandt, 1988). Some studies have also found that simulators are not as consistent in their performance (Brandt, 1988; Larrabee, 1991; Reitan & Wolfson, 1997; Strauss et al., 1999; Strauss et al., in press; Wasyliw & Cavanaugh, 1989).

Furthermore, one relatively consistent finding has been that simulators tend to do more poorly than head injured individuals on attention tasks. Specifically, on the Wechsler Memory Scale, simulators have been found to have reduced attention scores relative to their overall memory performance (Bernard, 1990; Boyer, 1991; Crossen & Wiens, 1988; Mittenberg et al., 1993; Reid & Kelly, 1991). Similarly, simulators tend to do more poorly on Digit Span than Vocabulary (WAIS-R), a discrepancy in the opposite direction than that observed in head injured individuals (Faust et al., 1988; Thompson & Cullum, 1991; Trueblood, 1994; Binder, 1990; Mittenberg et al., 1995). In addition, their performance on Reliable Digit Span (that level at which both digit span trials are passed) tends to be suppressed (Greiffenstein et al., 1996)

Overall, simulators have been found to have suppressed performance in comparison to head injured and/or normal controls on the following subtests of the WAIS-R and WMS:

- (1) Digit Span (Bernard, 1990; Binder & Willis, 1991; Heaton et al., 1978; Martin &

Franzen, 1992; Mittenberg et al., 1993; Rawlings & Brooks, 1990; Thompson & Cullum, 1991; Trueblood, 1994; Trueblood & Schmidt, 1993).

- (2) Orientation (Wiggins & Brandt, 1988).
- (3) Mental Control (Mitternberg et al., 1993).
- (4) Information (Bernard, 1990).
- (5) Logical Memory (Bernard 1990; Bernard et al., 1993).
- (6) Verbal Paired Associates (Bernard, 1990; Bernard, et al., 1993; Gronwall, 1991; Mittenberg et al., 1995).
- (7) Spatial Span (Bernard, 1990; Bernard et al., 1993; Mittenberg et al., 1993).
- (8) Digit Symbol (Trueblood, 1994; Trueblood & Schmidt, 1993)
- (9) Picture Completion (Trueblood, 1994; Trueblood & Schmidt, 1993)
- (10) Vocabulary (Trueblood & Schmidt, 1993).

Although there has been research (see above) exploring malingeringers' patterns of performance on the old version of the Wechsler scales (WAIS-R and WMS), some methodological concerns have been raised (see Rogers, 1997 for review). Table 2 lists simulation studies done to date, and some of the problems noted.

Of particular concern, the instructions to malingeringers have often been vague, not providing the subject with information that clinical malingeringers are likely to have. In addition, subjects have typically been presented with their instructions shortly before testing, thereby giving them little time to consider common post-injury symptoms or malingering strategies. It has been suggested that preparation time may be an important variable to manipulate, in order to more closely approximate real life malingering.

situations (Faust & Ackley, 1998), and that it may have an impact on subjects' level of motivation and participation (e.g., Rogers et al., 1993).

Table 2

Problems With Simulation Studies Using the Wechsler Scales

<u>Authors</u>	<u>Groups</u>	<u>Tests</u>	<u>Statistics</u>	<u>Problems</u>
Heaton et al., 1978	16 non-litig HI -mod-severe 16 Simulators	WAIS	T-test; DFA	Very small sample dropped subjects Exclusion criteria? Vague instructions Compliance check? Preparation time?
Hayward et al., 1987	21 LFT 28 Simulators -experienced nurses	Parts of WAIS & WMS	T-tests	Exclusion criteria? Vague instructions Preparation time? Heterogeneous HI -e.g., CVD
Bernard, 1990	28 NC 28 Simulators no incentive 30 Simulators with incentive	WMS-R	ANOVA DFA	Exclusion criteria? No HI group Vague instructions Cross validation? Preparation time?
Bernard et al., 1993	44 HI 89 Simulators	WMS-R	DFA	Exclusion criteria? Vague instructions Cross-validation? Preparation time? HI severity? Litigation status?
Mittenberg et al., 1993	39 non-litig HI -mild to severe 39 Simulators	WMS-R	T-Test; DFA	Exclusion criteria? Vague instructions Compliance check? Preparation time?
Mittenberg et al., 1995	67 non-litig HI -mild to severe 67 Simulators	WAIS-R	DFA	Vague instructions Compliance check? Exclusion criteria?
Johnson et al., 1998	15 warned simulators 15 unwarned simulators 15 NC	WAIS-R	MANOVA	Small Sample size No HI Group Compliance check?

Note. CVD = Cerebrovascular Disease; DFA= Discriminant Function Analysis; HI = Head Injury; LFT = Left Frontotemporal; NC = Normal Controls

Half of the research studies listed in Table 2 did not adequately assess the subjects' compliance. Some studies did not address this issue at all, while others assumed compliance because subjects appeared to answer symptom checklists in accordance with their instructions (e.g., Mittenberg et al., 1995). However, this is not the same as directly asking whether or not the subject followed instructions on the tests of interest. This seems particularly important given that some researchers have found that 10-18% of their sample did not comply with the instructions that they were given (e.g., Rose et al., 1998).

Haines and Norris (1995) also point out that many studies investigating malingering have failed to provide sufficient information regarding their simulating and control groups. A number of studies conducted do not report having screened for factors (e.g., history of neurological problems; drug use) that may have affected the test results of their control groups.

There have also been a number of problems with some of the head-injured samples employed. Some researchers (e.g., Hayward et al., 1987) have used individuals with neurological conditions (e.g. cerebrovascular disease) that can have a significantly different impact on cognitive ability than head trauma sustained through a motor vehicle accident or other external cause of injury. As these individuals do not present with the typical etiology of those who might be suspected of malingering, it would seem more reasonable to exclude them from the comparison groups.

In summary, to date the research on malingeringers' pattern of performance on test batteries has been far from adequate, and the findings have not been consistent (Rogers et al., 1993). In addition, although some information is available on malingeringers' performance on the old version of the Wechsler scales, a recent literature review revealed

only one analogue study pertaining to performance patterns on the new Wechsler Memory Scale-III. Specifically, a study by Killgore and DellaPietra (2000) revealed that there are 6 items on the recognition portion of the WMS-III Logical Memory subtest that are rarely missed, even by individuals naïve to the content of the stories. These researchers found that a weighted combination of these six items accurately classified 98% of participants and demonstrated high sensitivity (97%) and specificity (100%) in discriminating between analogue malingeringers and patients. However, information on the pattern of performance of simulators on other subtests of the new Wechsler scales remains limited.

In addition, there is some evidence that the assessment of response consistency may prove to be an especially powerful tool for the assessment of malingering (e.g., Cullum, Heaton, & Grant, 1991; Grote et al., 2000; Strauss et al., 1999 and in press). Wetter and Deitsch (1996) evaluated the ability of college students to consistently fake symptoms of PTSD and CHI. Although participants were able to consistently fake symptoms of PTSD upon retesting two weeks later, there was considerable variability in the performance of simulators faking symptoms of closed head injury.

Further support for the potential usefulness of the assessment of consistency comes from the work of Reitan and Wolfson (1995, 1996, 1997) which revealed that the performance of head injured individuals involved in litigation was less consistent than the performance of non-litigating patients. Recent work by Strauss and colleagues (1999) revealed evidence of substantial inconsistency in the cognitive performance of malingeringers, especially on tests of symptom validity. It is important to note, however, that inconsistency has been linked to bona fide central nervous system problems (Strauss et al., 1999). Little research has explored consistency of performance on the new Wechsler scales.

Proposed study

The current study assessed the pattern of malingered performance on the WAIS-III and WMS-III, as well as on a self-report measure (i.e., British Columbia Postconcussion Symptom Inventory) of physical and psychological symptoms often reported following traumatic brain injury. The study was conducted using a simulation paradigm, in which normal control subjects were asked to fake symptoms of a mild head injury (Simulators). The performance of the simulators was then compared to that of a normal control group and a group of individuals seen for evaluation of a possible head injury. This study also aimed to take into consideration some of the methodological concerns detected in previous studies.

Methodological Considerations

Attempts were made to address the following methodological problems that have been noted in earlier studies (see Table 2):

- (1) Unlike Heaton and colleagues (1978) who had a small sample, the proposed study has a sample size that supports the statistical analyses considered.
- (2) Participants were screened for factors that could influence test results, but were not related to the question of interest (e.g., drug and alcohol problems, previous HI).
- (3) The simulators were provided with instructions in advance of the testing session, thereby providing them with the opportunity to think about sequelae of head injury and malingering techniques.

- (4) The instructions were not as vague as those used in previous research (see Klimczak, Donovick, & Burright, 1997), providing the participants with some indication of symptoms that could be experienced following a head trauma.
- (5) Following completion of the testing session, participants were asked whether or not they were able to comply with the test instructions.
- (6) The head injury group only included individuals who sustained a head injury due to external factors (e.g., motor vehicle accidents, falls). An attempt was made to restrict the group to people who have sustained a motor vehicle accident.

Hypothesis I: Subjective Complaints

A self-report measure (e.g. BCPSI) was employed to assess physical and psychological symptoms often reported following head injury (i.e., postconcussive syndrome). It was anticipated that like their performance on cognitive tasks, simulators would exaggerate their deficits and would thus endorse more items than the normal controls.

Hypothesis II: Performance on WAIS-III and WMS-III

Based on research findings reported to date, one of two possible patterns of performance was expected on the cognitive measures.

- 1) In line with the findings of some researchers (e.g., Klimczak et al., 1997), simulators might suppress their performance on most tasks, perhaps to a greater extent than that typically observed in truly head injured individuals (e.g., Trueblood 1994). This general suppression of performance could possibly be explained by laypersons'

common misconceptions regarding head injury (e.g., loss of fund of knowledge, loss of knowledge regarding personal history).

- 2) Alternatively, as found in previous research, the simulators could present with specific areas of poor performance. Specifically:
 - A) In keeping with the pattern typically observed in brain injured individuals, simulators could be expected to score more poorly on the non-verbal component (Performance Intelligence Quotient, PIQ) than on the verbal component (Verbal Intelligence Quotient, VIQ) of the Wechsler intelligence scale (e.g., Mittenberg et al., 1995; Rawlings & Crewe, 1992).
 - B) In addition, the Simulators could perform more poorly than normal controls or litigants on tasks with an attentional component. Previous studies, (Bernard, 1990, 1991; Boyer, 1991; Iverson et al., 2000; Johnson & Lesniak-Karpiak, 1997; Mittenberg et al., 1993; Reid & Kelly, 1991), have shown that on the WMS, simulators especially suppress their performance on attention/ working memory measures relative to general memory (e.g., discrepancy between General Memory Index and Attention/Concentration Index). Similar results in discrepancy between WMS-III General Memory and Working Memory Index (WMI) were expected, given that the WMI and Attention/Concentration index correlate moderately well (e.g., .64-.73) and have both been shown to have good discriminant validity (The Psychological Corporation, 1997).

With regard to WAIS-R performance, reduced performance on attention tasks has been noted in the evaluation of Voc-DS discrepancy scores. Specifically, while head injured individuals do not typically show significant discrepancy in their performance on these tasks (e.g., Mittenberg et al., 1995), simulators have

been found to suppress their performance on the Digit Span subtest, resulting in a larger discrepancy score (e.g., Mittenberg et al., 1995; Strauss et al., 1999; Trueblood, 1994; Thompson & Cullum, 1991).

In addition, evidence of suppressed performance on attention measures has been obtained through the evaluation of a derived score based on Digit Span performance (e.g., Reliable Digit Span; Greiffenstein et al., 1994; Strauss et al., 1999). Reliable Digit Span is calculated by summing the longest string of digits repeated without error over two trials, under both forward and backward conditions. It was hypothesized that simulators might have a lower Reliable Digit Span than the other groups, and similarly might perform more poorly on the corresponding derived score for the Spatial Span subtest (e.g., Reliable Spatial Span).

Hypothesis III: Comparison of WAIS-III and K-BIT Performance

The effect of biased responding on overall intellectual functioning was also explored by comparing the simulators' scores on the WAIS-III (less than optimal effort put forth) to those obtained with optimal effort on the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990). Previous research has shown that biased responding results in lower overall IQ, as determined by comparing malingered WAIS IQ scores to estimates of actual IQ (e.g., Heaton et al., 1978). No studies to date, however, have compared malingered versus actual IQ within subjects on standardized tests of intelligence. It was hypothesized that the SIM group would not demonstrate the same pattern of performance as the NC group.

Hypothesis IV: Consistency of Performance

Consistency of performance was also examined by comparing group performances on the Digit Span and Letter-Number Sequencing subtests of the WAIS-III, to group performances on these same tasks administered as part of the WMS-III a few hours later. As noted by a number of researchers (e.g., Miller & Miller, 1992; Reitan & Wolfson, 1997; Strauss et al., 1999), response consistency may be one of the most powerful methods available to neuropsychologists for detecting invalid test results. It was hypothesized that the simulators would have a greater difference in scores from one administration to the next.

Hypothesis V: 21-Item Test

It was hypothesized that the simulators would perform more poorly than the NC group on the 21-Item Test, a task specifically designed to detect malingering (Iverson et al., 1994).

Method

Participants

Three target samples were recruited for this study: a group of healthy individuals asked to simulate the effects of mild head injury (SIM), a litigating group (LIT), and a normal control group (NC).

The SIM and NC groups were drawn from the undergraduate population at the University of Victoria. Seventy-two participants were obtained through the Psychology 100 Research Pool and received extra credit points for their participation. Subjects were excluded ($n = 1$) if they were non-native English speakers, or had a past or present history of neurological, psychiatric, or drug and alcohol problems as assessed through self-report. Data from five participants were excluded from the analyses because they did not follow instructions as assessed through the post-testing compliance check. The resulting normal control group (NC) consisted of 34 participants who averaged 23.1 years of age ($SD = 5.9$, range 17-43), with a mean education level of 14.5 years ($SD = 2.1$). The simulating group (SIM) consisted of 32 participants who averaged 23.4 years of age ($SD = 6.7$, range 18-48), with a mean education level of 14.4 years ($SD = 1.1$).

The litigating group (LIT) was obtained through the private practice offices of Drs. King and Strauss, Psychologists. In total, data on 34 individuals reporting cognitive deficits following the experience of a head injury were obtained. The litigating individuals were classified by the consulting psychologist as having sustained a mild, moderate, or severe head injury based on the following severity indices: Glasgow Coma Scale (GSC), loss of consciousness (LOC), post-traumatic amnesia (PTA) and neuroimaging findings.

The litigating groups consisted of 22 litigants reporting a history of mild head injury (LIT), 8 litigants classified as moderately to severely head injured individuals (MSLIT), and 4 suspect malingeringers (SUSP). The LIT group averaged 32.7 years of age (SD = 9.4, range 18-48) with 12.4 years of education (SD = 1.7), and had sustained their injuries as a result of a motor vehicle accident (59.1%), motorcycle accident (13.6%), fall (13.6%), assault (9.1%), or cycling accident (4.5 %). The LIT individuals were tested, on average, 2.3 (SD = 2.0, range 0.17-8.0) years after sustaining the head injury. All individuals in the LIT group reported cognitive symptoms and only two were tested within three months of their injury. The moderate to severe head injury litigants (MSLIT) were not considered in the statistical analyses of this study due to the small group size (n = 8). However, information on their performance is provided for casual comparison. Similarly, information is provided on four head-injured individuals (three LIT, one MSLIT) who, for the purpose of this study, were classified as having put forth "suspect" effort (SUSP) based on their performance on at least one test specifically assessing biased responding: VSVT (Slick, Hopp, Strauss, & Thompson, 1997), TOMM (Tombaugh, 1996), and the 21-Item Test (Iverson et al., 1991, 1994).

A compromise power analysis (Erdfelder, Faul, & Buchner, 1996) was conducted to estimate power for the present investigation. Compromise power analyses are particularly important when working with clinical populations whose constraints (i.e., small n) make conventional a priori and post hoc power analyses inappropriate (Erdfelder et al., 1996). Compromise power analyses take into account the relative trade-off between demands for low alpha and large power levels. The results of the analysis for three groups (n = 88), assuming a medium effect size (.25) and a Beta/Alpha ratio of 1, yielded an estimated power = .79.

One-way analysis of variance was used to examine group differences on the demographic variables of age and education; the main effect of group was significant for both age, $F(2, 84) = 13.24, p < .001, \eta^2 = .24$ and for education $F(2, 84) = 11.67, \eta^2 = .22$. Post-hoc comparisons were examined using Tukey's HSD procedure at the $\alpha = .05$ level for all pairwise comparisons. There were no significant differences between the NC and SIM groups for either age or education. Both groups, however, were significantly different from the litigants, with the LIT group being significantly older and less educated than the NC and SIM groups. See Table 3 for demographic information on the groups.

Table 3

Demographic Information on NC, SIM, and Litigating Groups (LIT, MSLIT, & SUSP)

Demographics	Groups				
	NC	SIM	LIT	MSLIT	SUSP
Sample Size	34	32	22	8	4
Age	23.1	23.4	32.7	36.0	37.4
SD	5.9	6.7	9.4	12.4	12.4
Age Range	18-42	18-48	19-48	20-50	21-50
Education	14.5	14.4	12.4	12.9	13.3
SD	2.1	1.1	1.7	2.0	2.2
Gender %	67.6 F 32.4 M	59.4 F 40.6 M	31.8 F 68.2 M	62.5 F 37.5 M	75 F 25 M
Handedness %	91.2 R 8.8 L	87.5 R 12.5 L	81.8 R 18.2 L	100 R 0 L	75 R 25 L

Note. NC = Normal Controls, SIM = Simulators, LIT = Mildly head-injured litigants, MSLIT = Moderately to severely head-injured litigants, SUSP = Suspect malingeringers.

Design and Procedure

Participants from the NC group and SIM group were asked to retrieve a file from the Psychology Department one to two days before testing. Failure to do so resulted in the cancellation of the testing session. The file they retrieved contained two numbered envelopes to be opened and completed in the specified order. The first envelope included a questionnaire regarding personal and demographic information such as age, gender, level of education, as well as neurological, psychiatric, and general health history. The second envelope contained printed instructions for the assessment, and the BCPSI symptom checklist (i.e., physical and psychological symptoms) to be completed according to the instructions.

The instructions for the control group requested that they put forth their best effort on the tasks to be completed, while instructions for the simulating group asked them to realistically fake symptoms of a mild head injury, but avoid detection. The text of these instructions read as follows:

Control Group

Thank you for taking part in this experiment. You will be completing a variety of different tests and procedures. Some of the tasks will just involve you answering questions (e.g., giving word definitions) and others will involve more 'hands-on' activities (e.g., arranging blocks). You will find that the tasks typically start off easy and then get harder. Most people don't answer every question correctly or finish every item. The important thing is that you give your best effort on all the items. You will find that some of the tasks will be easy for you and that others may be more difficult. Just try your best.

You are part of the control group for this experiment. The person testing you is a "blind experimenter." In other words, she has not been told the purpose of the study, in order to assure that she tests the control group and experimental group in the same manner. It is important that you do not talk to her about the purpose of the study or the group that you are in, at any point during the experiment.

Simulating Group

I would like you to pretend that 6 months ago, you were involved in a car accident. You hit your head on the dashboard and you were knocked out (lost consciousness) for approximately 5

minutes. The accident was not your fault. You were taken to the hospital by ambulance, observed for a few hours, and then released. You were told that in addition to sustaining a concussion, you had soft tissue damage in your neck. Brain scans done at the hospital were normal and the doctor told you that you could expect a full recovery.

You were off work for several weeks due to back and neck pain. When you returned home following the accident, you were confused, could not remember what had happened, and felt that you "couldn't think straight". You were forgetful and had difficulty concentrating. You also felt irritable, "on edge", and you had daily headaches.

Over time, your condition improved greatly. However, you are now feeling quite anxious about the lawsuit you initiated to get compensation for the injuries you sustained in the accident. You have undergone numerous evaluations and feel that people do not believe that there was, or is anything wrong with you as a result of the accident.

You are about to see a psychologist who is going to give you a battery of tests designed to measure your concentration, memory, and other thinking skills. You have decided to exaggerate on the tests, to make your symptoms look worse than they are. Your goal is to appear realistically impaired on these tests. It is important that your performance on these tests does not make it obvious that you are faking, as this would result in the loss of the settlement you are trying to obtain.

The person testing you is a "blind experimenter". In other words, she has not been told the purpose of the study, in order to assure that she tests the control group and experimental group in the same manner. It is important that you do not talk to her about the purpose of the study or the group that you are in, at any point during the experiment.

No other information about head injury symptoms or test procedures was provided. Although the experimenter did have knowledge regarding the nature of the study, she was unaware of the subject's group status until after the testing session. Subjects were lead to believe that they were being assessed by an experimenter who was not aware of the purpose of the study, to dissuade them from talking about the experiment during the evaluation.

Prior to starting the testing session, subjects were asked to review the content of their envelopes. Each control and simulating subject was then tested for approximately four hours, during which time they signed a consent form, and completed the WAIS-III (test 1-13 (Wechsler, 1997a); the WMS-III (all but the Visual Reproductions subtest; Wechsler, 1997b), the 21-Item Test (Iverson et al., 1994), and the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990) according to the standardized administration procedures. The subjects completed the WAIS-III, WMS-III and 21-Item test according to the written instructions they were given. However, prior to completing

the K-BIT, which was administered last, the participants were requested to disregard the instructions they had been given for the assessment and to put forth their best effort.

Following completion of the testing session, and after being reassured that their answers would not affect their receiving class participation points, subjects were asked a number of questions to evaluate compliance with the test instructions. Participants were asked when they opened the envelopes, read the instructions, and completed the questionnaires. No one reported less than 24 hours between reading the instructions and the testing session. Participants were also asked if they were able to comply with the written instructions that they received. Lastly, subjects were asked if they put forth their best effort on the K-BIT. All simulators included in the analyses reported disregarding the malingering instructions prior to completing the K-BIT. Participants then received information regarding the study and were given an opportunity to ask questions.

The litigants received the WMS-III and WAIS-III as part of a more extensive neuropsychological evaluation. They did not complete the K-BIT, 21-Item Tests, or any of the questionnaires (e.g., BCPSI), and thus could not be included in the analyses of these measures.

Materials

BCPSI. The British Columbia Postconcussion Symptom Inventory is an unpublished self-report questionnaire created by Grant Iverson (1998), which assesses the frequency, intensity, and duration of physical symptoms, as well as the frequency and intensity of psychological symptoms commonly reported following head injury.

21 Item Test. The 21-Item Test (Iverson et al., 1994) is a 21 item word list that, upon being read to the subject, is followed by a free-recall component and a two-alternative, forced choice recognition procedure. In the recognition procedure, each word from the original list is paired with a distractor word. Of the 21 word pairs, seven are semantically related, seven pairs rhyme, and seven do neither. The general notion is that recognition memory should always be better than free recall, and that recognition memory should be superior to chance. A cutoff score of 13 is suggested on the forced-choice procedure as indicating questionable performance (Iverson & Franzen, 1996). In addition to performance on the recognition portion of the task, two other scores from the 21-item test are thought to be promising indicators of biased responding. The inconsistency score corresponds to the number of items remembered on the free recall portion of the test that are then missed on the recognition task. Normal controls and patients putting forth good effort do not typically have such inconsistencies. The consecutive misses score corresponds to the largest number of consecutive misses on the recognition task. A large number of consecutive misses is also considered unlikely in individuals putting forth good effort (Spreen & Strauss, 1998).

K-BIT. The Kaufman Brief Intelligence Test (Kaufman & Kaufman, 1990) is a measure that provides a brief estimate of intelligence. It consists of two subtests, Matrices and Vocabulary, which provide an estimate of nonverbal IQ and verbal IQ respectively. The Matrices subtest is comprised of both concrete and abstract stimuli. The Vocabulary subtest is subdivided into two parts: the Expressive component requires the subject to provide the name for a pictured object presented, while the Definitions component requires the subject to provide the word that best fits verbal and visual clues

(i.e., verbal description; partially complete word). The combined scores from the Vocabulary and Matrices subtests provide a Composite IQ.

The K-BIT Composite score has been found to correlate moderately well with other measures of intelligence such as the WAIS-R (FSIQ) (see Spreen & Strauss, 1998 for review), with the Composite score tending to be about five points higher than the WAIS-R FSIQ. However, the subcomponents of the KBIT (Matrices and Vocabulary) only correlate modestly with the verbal and non-verbal components of the WAIS-R. No research to date has explored the relation of the K-BIT to the WAIS-III.

WAIS-III (Psychological Corp., 1997a). The Wechsler Adult Intelligence Scale-Third Edition yields the traditional three composite IQ scores (FSIQ, VIQ, PIQ), as well as four index scores (Verbal Comprehension; Perceptual Organization; Working Memory; Processing Speed). This test is comprised of 14 subtests each measuring a facet of intelligence. Of the 14 subtests, three are new to this edition:

- (1) Symbol Search: A measure of attention and concentration, visual perception, spatial orientation, and speed.
- (2) Matrix Reasoning: A measure of visual information processing and abstract reasoning skills, thought to enhance the assessment of nonverbal, fluid reasoning.
- (3) Letter-Number Sequencing: Designed as a measure of attention and working memory, that uses auditory stimuli.

WMS-III (Psychological Corp., 1997b). The Wechsler Memory Scale- Third Edition provides eight summary index scores: Auditory Immediate Index, Visual

Immediate Index, Immediate Memory Index, Auditory Delayed Index, Visual Delayed Index, Auditory Recognition Delayed Index, Working Memory Index, and General Memory Index. This measure consists of 11 subtests assessing learning, as well as immediate, delayed and working memory. Each of these domains is tested in the auditory and visual modalities. Of the 11 subtests, four are new to this edition:

- (1) Faces: Measures immediate and delayed memory using a recognition paradigm.
- (2) Family Pictures: Measures ability to recall complex and meaningful, visual information. This subtest is designed to assess recall for characters, spatial location, and scene activity.
- (3) Word Lists: List learning paradigm measuring immediate and delayed memory.
- (4) Letter-Number-Sequencing: Measure of attention and working memory, which uses auditory stimuli.

Results

To control for the violation of Type I experiment-wise error rate, a more conservative $\alpha = .01$ was adopted to interpret significant results at the univariate level. Conservative control of Type I error at the expense of Type II error was employed to avoid inaccurate classification of individuals as biased responders, given the stigma and negative consequences associated with this label. All follow-up post hoc comparisons were conducted using Tukey's HSD procedure, Bonferroni adjustments, or Student-Newman-Keul at the $\alpha = .05$ family-wise level.

The results will be presented in the following order, comparing the groups (NC, SIM, LIT) on:

- 1) Subjective Complaints
- 2) WAIS-III Summary Scores
- 3) WAIS-III Discrepancy Scores
- 4) WAIS-III Subtest Scores
- 5) Comparison of WAIS-III and K-BIT Performance
- 6) WMS-III Summary Scores
- 7) WMS-III GMI-WMI Discrepancy Score
- 8) WMS-III Subtest Scores
- 9) Reliable Digit Span and Reliable Spatial Span
- 10) Consistency of Performance
- 11) 21-Item Test

As mentioned in the Methods section, information on a moderate to severe head injury litigating group (MSLIT) and litigants thought to be putting forth suspect

performance (SUSP) will be provided for casual comparisons. However, these two groups were not considered in the formal statistical analyses due to the small group sizes ($n = 8$ and 4 respectively).

Part I: Subjective Complaints

The BCPSI Questionnaire was employed to examine group (NC and SIM) differences for self-reported physical and psychological symptoms. Unfortunately, this questionnaire was changed part way through the experiment, resulting in smaller group sizes (i.e., NC = 15, SIM = 14). Comparisons with the LIT group were not possible, as this measure was not administered in the private practices from which the litigants were derived.

Overall, it was hypothesized that the SIM group would have more subjective complaints than the NC group. A multivariate analysis of variance (MANOVA) was used to compare the SIM and NC groups on the occurrence, intensity and duration of a number of physical and psychological symptoms commonly reported following head injury (see Mittenberg et al., 1995). The omnibus effect for group was significant, $F(5, 23) = 3.16, p = .03, \eta^2 = .41$. The SIM group's experience of symptoms was significantly different (e.g., higher scores) than that reported by the NC group, suggesting an understanding of and compliance with instructions. Although univariate analyses failed to satisfy the $\alpha = .01$ criterion, the mean trends (see Table 4) were all in the expected direction. In particular, univariate analyses approached significance for intensity of physical symptoms ($F(1, 27) = 5.69, p = .02, \eta^2 = .17$) as well as frequency ($F(1, 27) = 5.38, p = .03, \eta^2 = .17$) and intensity of psychological symptoms ($F(1, 27) = 6.51, p = .02, \eta^2 = .19$).

Table 4

BCPSI Summary Scores

Symptoms	Group			
	Normal Control (n = 15)		Simulators (n = 14)	
	<u>M</u>	(SD)	<u>M</u>	(SD)
Physical				
Frequency	19.60	(12.86)	27.36	(15.23)
Intensity*	14.93	(9.18)	25.14	(13.59)
Duration	20.93	(11.49)	26.93	(14.74)
Psychological				
Frequency*	21.13	(15.35)	36.29	(19.71)
Intensity*	21.10	(15.29)	37.43	(19.15)

* $p < .05$

Multivariate analysis of variance was also used to compare the SIM and NC groups' performance on specific items from the BCPSI measure, reflecting common symptoms reported after mild head injury (see Table 5)¹. The omnibus effect for group was significant, $F(23, 39) = 8.04, p < .01, \eta^2 = .83$. With regard to physical symptoms, at the univariate level, the simulators reported experiencing more frequent ($F(1, 61) = 26.74, p < .01, \eta^2 = .31$) and more intense ($F(1, 61) = 24.39, p < .01, \eta^2 = .29$) headaches, as well as more frequent ($F(1, 61) = 20.55, p < .01, \eta^2 = .25$) and more intense ($F(1, 61) = 21.88, p < .01, \eta^2 = .26$) dizzy spells than the normal controls. In addition, the SIM group experienced more intense fatigue ($F(1, 61) = 9.85, p < .01, \eta^2 = .14$) and more intense blurred vision ($F(1, 61) = 8.04, p < .01, \eta^2 = .12$). Sleep disturbances were also more frequent ($F(1, 61) = 23.67, p < .01, \eta^2 = .28$), intense ($F(1, 61) = 18.06, p < .01, \eta^2 = .23$), and long lasting ($F(1, 61) = 9.73, p < .01, \eta^2 = .14$) in the SIM group.

Table 5

BCPSI Psychological and Physical Symptoms Reported by the NC and SIM Groups

	Group			
	Normal Controls (n = 32)		Simulators (n = 31)	
	M	(SD)	M	(SD)
Physical Symptoms				
Headache F**	2.19	(1.31)	3.74	(1.03)
Headache I**	2.09	(.93)	3.13	(.72)
Headache D	3.28	(.92)	3.65	(.88)
Dizziness F**	1.13	(1.18)	2.58	(1.36)
Dizziness I**	.97	(1.06)	2.39	(1.33)
Dizziness D	1.09	(1.20)	1.38	(.80)
Blurred Vision F*	.41	(1.13)	1.16	(1.55)
Blurred Vision I**	.31	(.82)	1.26	(1.69)
Blurred Vision D	.47	(1.27)	1.06	(1.48)
Sleep ProblemsF**	1.50	(1.46)	3.32	(1.51)
Sleep ProblemsI**	1.44	(1.50)	3.00	(1.41)
Sleep ProblemsD**	2.28	(1.89)	3.65	(1.56)
Fatigue F	2.59	(1.58)	3.35	(1.74)
Fatigue I**	1.91	(1.20)	3.00	(1.55)
Fatigue D	3.44	(1.32)	3.84	(1.59)
Psychological Symptoms				
Memory F**	.72	(1.33)	3.35	(1.43)
Memory I**	.59	(1.21)	3.45	(1.34)
Concentration F**	1.47	(1.29)	3.81	(1.05)
Concentration I**	1.59	(1.24)	3.74	(1.09)
Irritability F**	1.53	(1.08)	3.42	(1.20)
Irritability I**	1.59	(1.16)	3.16	(1.07)
Sadness F	1.28	(1.35)	1.97	(1.58)
Sadness I*	1.28	(1.39)	2.23	(1.71)

Note: F = Frequency, I = Intensity, D = Duration.

* p < .05. ** p < .01.

With regard to psychological symptoms, the univariate analyses revealed that the SIM group endorsed more frequent ($F(1, 61) = 57.80, p < .01, \eta^2 = .49$) and intense ($F(1, 61) = 78.93, p < .01, \eta^2 = .56$) memory problems, more frequent ($F(1, 61) = 61.89, p < .01, \eta^2 = .50$) and intense ($F(1, 61) = 53.00, p < .01, \eta^2 = .47$) concentration problems, and more frequent ($F(1, 61) = 43.06, p < .01, \eta^2 = .41$) and intense ($F(1, 61) = 31.10, p < .01, \eta^2 = .34$) irritability. Based on our $\alpha = .01$ criterion, the SIM and NC groups did not differ in their endorsement of sadness.

Part II: WAIS-III Summary Scores

Means and standard deviations for the WAIS-III summary scores are presented in Table 6 and are graphically displayed in Figure 1. One of two possible outcomes was expected. In keeping with previous research findings (e.g., Trueblood, 1994), simulators might have suppressed their overall performance, resulting in lower summary scores than observed in the other groups. Alternatively, the simulators could present with specific areas of poor performance discernable, in the evaluation of discrepancy scores (see Part III).

A profile analysis controlling for the effects of education was conducted to examine group (NC, SIM, LIT) differences on the Verbal Intelligence Quotient (VIQ), Performance Intelligence Quotient (PIQ), Full Scale Intelligence Quotient (FSIQ), Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Working Memory Index (WMI), and Processing Speed Index (PSI) shown in Figure 1. Only the effect of parallelism was of interest, as it provides information regarding the similarity of profiles between the groups (i.e., the group by task interaction)².

Table 6

Means and Standard Deviations for WAIS-III Summary Scores and Discrepancy Scores

Summary Scores	Group				
	NC (n = 34)	SIM (n = 32)	LIT (n = 22)	MSLIT (n = 8)	SUSP (n = 4)
	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>
FSIQ*	111.68 (8.66)	106.41 (10.15)	104.05 (12.52)	99.29 (12.35)	94.00 (9.56)
VIQ*	113.12 (9.61)	109.28 (11.32)	102.23 (11.72)	99.29 (11.04)	95.00 (4.08)
PIQ*	108.41 (10.45)	101.78 (10.54)	106.00 (14.81)	99.50 (13.15)	93.00 (15.58)
VCI*	114.53 (12.42)	110.91 (11.99)	103.00 (11.11)	97.50 (9.32)	94.50 (9.33)
POI	107.59 (11.30)	104.69 (11.08)	109.77 (16.19)	101.38 (15.37)	96.50 (16.09)
WMI	104.65 (9.08)	101.09 (12.19)	100.00 (13.51)	101.14 (14.99)	95.75 (15.28)
PSI*	109.47 (13.65)	93.88 (13.74)	97.64 (11.90)	96.88 (16.33)	81.75 (10.24)
VIQ-PIQ	4.71 (12.76)	7.38 (10.84)	-3.77 (11.94)	-.29 (10.98)	2.00 (11.63)
VOC-DS*	3.38 (3.11)	3.00 (3.38)	1.18 (2.89)	-.63 (3.34)	-.25 (5.85)

Note. FSIQ = Full Scale Intelligence Quotient, VIQ = Verbal Intelligence Quotient, PIQ = Performance Intelligence Quotient, VCI = Verbal Comprehension Index, POI = Perceptual Organization Index, WMI = Working Memory Index, PSI = Processing Speed Index, VOC = Vocabulary, DS = Digit Span.

*p < .05. **p < .01. for comparisons among NC, SIM, and LIT groups.

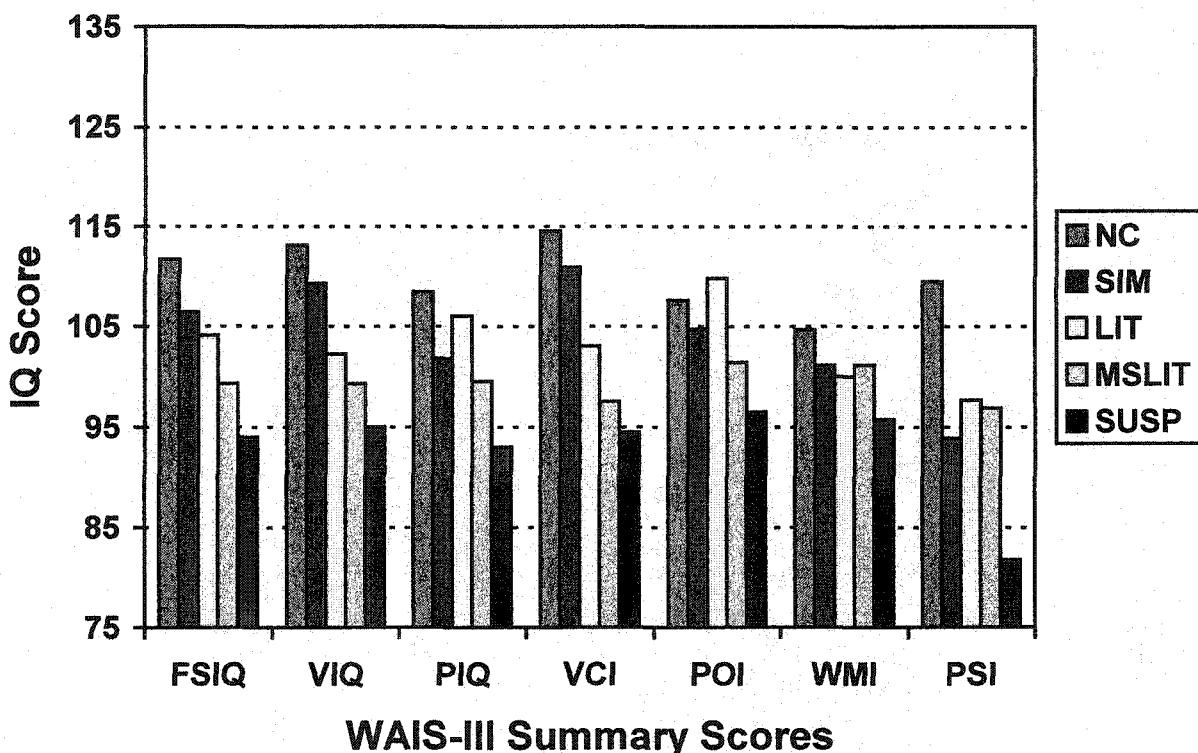


Figure 1. Group Profiles on WAIS-III Summary Scores.

The test of parallelism revealed a significant interaction, $F(12, 504) = 3.86, p < .001, \eta^2 = .08$. For each dependent measure, post-hoc comparisons (Student-Newman-Keul) were conducted to examine all pairwise group comparisons. All reported comparisons were significant $\alpha = .05$.

The LIT group performed significantly more poorly than the NC group on FSIQ, VIQ, VCI, and PSI. The SIM group's performance was lower than the NC group's performance only on PIQ and PSI. The SIM group and LIT group only differed significantly on the VIQ and VCI, with the LIT group performing significantly worse than the SIM group.

Overall, the performance of the SIM group tended to be suppressed, but generally not enough to differ significantly from either the NC or LIT groups. There was a non-significant trend for the SIM group to perform even more poorly than the LIT group on PIQ, POI, and PSI indices, thus more closely resembling the performance of moderate to severely head injured litigants (MSLIT).

Although statistical comparison with the MSLIT and SUSP groups was conducted, visual inspection suggests that the SIM group performed more poorly than the MSLIT group on PSI. Interestingly, the mildly head-injured litigants thought to be putting forth less than optimal effort (SUSP) appeared to consistently perform more poorly than all groups, including the MSLIT group.

Part III: WAIS-III Discrepancy Scores

Although some researchers (e.g., Klimczak et al., 1997) have reported overall suppressed scores, specific areas of poor performance have been reported in other research (e.g., Mittenberg et al., 1995). With regard to WAIS-III performance, a one-way analysis of covariance (ANCOVA) controlling for education was computed to compare the NC, SIM, and LIT groups on the discrepancy between VIQ and PIQ scores (see bottom of Table 6). In keeping with previous research findings, it was expected that the SIM group would score more poorly on PIQ than on VIQ, a pattern similar to that seen in head injured individuals (e.g., Mittenberg et al., 1995). Results revealed, however, that the univariate comparison of the three groups on VIQ-PIQ discrepancy was not significant.

The NC, SIM, and LIT groups were also compared on the Vocabulary-Digit Span Discrepancy score. This derived score has been found useful in detecting malingered performance (e.g., Mittenberg et al., 1995). In keeping with previous research findings, it was expected that the SIM group would suppress their performance on Digit Span in comparison to their performance on Vocabulary. The SIM group's discrepancy score was thus expected to be larger than that observed in the NC and LIT groups.

ANCOVA, controlling for education, was computed to compare the NC, SIM, and LIT groups on the discrepancy between Vocabulary and Digit Span scores (see bottom of Table 6). The univariate comparison of the three groups on Vocabulary- Digit Span Discrepancy approached the specified $\alpha = .01$ criterion, $F(2, 84) = 3.89$, $p = .02$, $\eta^2 = .09$. Post-hoc contrasts revealed that only the LIT group significantly differed from the NC group. Interestingly, the NC group unexpectedly had the largest discrepancy scores.

Part IV: WAIS-III subtests

All WAIS-III subtests were considered in the analyses for exploratory purposes. Again, one of two patterns of performance was expected: overall suppression of performance or poor performance on tasks with an attentional component. Separate analyses were conducted on subtests with age-adjusted scaled scores and on subtests for which only raw scores were available.

MANCOVA controlling for education was used to examine group (NC, SIM, LIT) differences on WAIS-III age-adjusted scaled scores (see Table 7)³. The omnibus main effect of group was significant, $F(22, 144) = 2.80$, $p < .001$, $\eta^2 = .30$. Univariate group effects were significant for Digit Symbol, $F(2, 82) = 15.36$, $p < .01$, $\eta^2 = .27$, and

for Symbol Search, $F(2, 82) = 5.45, p < .01, \eta^2 = .12$ (two univariate effects approached the $\alpha < .01$ criterion: Vocabulary, $F(2, 82) = 4.54, p = .01, \eta^2 = .10$ and Matrix Reasoning, $F(2, 82) = 3.97, p = .02, \eta^2 = .09$). For both Digit Symbol and Symbol Search, post-hoc comparisons using Bonferroni adjustments revealed that the LIT and SIM groups performed significantly worse than the NC group. Although no task differentiated the SIM group from the LIT group, there was a non-significant trend for the SIM group to perform even more poorly than the LIT group on Matrix Reasoning and Symbol Search. Overall, the performance of the SIM group appeared to be suppressed but, once again, not enough to differ significantly from the NC or LIT groups.

Table 7

Means and Standard Deviations for WAIS-III Age-Adjusted Subtest Scores

Subtest	Group				
	NC (n = 34)	SIM (n = 30)	LIT (n = 22)	MSLIT (n = 8)	SUSP (n = 4)
	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>
Vocabulary*	13.12 (2.38)	12.23 (2.42)	10.95 (2.48)	9.75 (2.05)	9.00 (1.35)
Similarities	12.53 (2.56)	11.63 (2.91)	10.45 (1.90)	10.25 (2.19)	9.75 (1.03)
Arithmetic	12.06 (1.52)	11.40 (2.33)	10.50 (2.79)	9.86 (3.58)	10.25 (.48)
Digit Span	9.74 (2.33)	9.37 (3.03)	9.77 (2.67)	10.38 (2.72)	9.25 (2.02)

Table 7 con't.

Means and Standard Deviations for WAIS-III Age-Adjusted Subtest Scores

Subtest	Group				
	NC (n = 34)	SIM (n = 30)	LIT (n = 22)	MSLIT (n = 8)	SUSP (n = 4)
	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>	<u>M</u> <u>(SD)</u>
Information	12.24 (2.71)	11.77 (2.54)	10.41 (2.70)	8.88 (1.73)	8.25 (.85)
Comprehens.	13.35 (1.91)	12.50 (1.95)	-	11.20 (1.02)	-
L-N Sequen.	10.82 (2.29)	9.67 (2.29)	10.00 (2.78)	10.75 (1.03)	8.50 (1.71)
Picture Compl.	9.38 (2.66)	10.30 (2.22)	10.95 (3.30)	8.63 (.82)	8.00 (1.22)
Digit Symbol**	11.85 (2.60)	8.37 (2.57)	8.86 (3.11)	10.13 (1.20)	7.25 (1.11)
Block Design	12.41 (3.05)	12.43 (2.11)	11.95 (3.64)	11.00 (.96)	10.25 (1.49)
Matrix Reas*	12.21 (2.29)	10.43 (3.10)	11.86 (2.49)	11.25 (1.16)	10.25 (1.80)
Picture Arrang.	10.65 (2.46)	10.75 (3.20)	-	8.00 (.89)	-
Symbol Search**	11.59 (2.90)	9.70 (2.97)	10.41 (2.11)	8.75 (1.01)	6.00 (1.08)

Note. Missing data is represented by a dash (-).

Comprehens. = Comprehension, L-N Sequen. = Letter Number Sequencing, Picture Compl. = Picture Completion, Matrix Reas. = Matrix Reasoning, Picture Arrang. = Picture Arrangement.
* $p < .05$. ** $p < .01$. for comparisons among the NC, SIM, and LIT groups.

Due to missing data for the LIT group on Comprehension and Picture

Arrangement, a separate MANOVA was computed to examine group differences for the NC and SIM groups on these two subtests. The omnibus group effect was not significant and the univariate analyses were thus not explored.

A multivariate analysis of covariance (MANCOVA) controlling for the influence of age and education was conducted on WAIS-III subtests for which only raw scores were available (see Table 8). As the omnibus effect for group was not significant, the univariate effects were not considered.

Due to missing data for the LIT group on Digit Symbol Pairing and Free Recall, a separate MANOVA was computed to examine group differences for the NC and SIM groups on these two variables. The omnibus group effect was significant, $F(2, 62) = 15.78, p < .01, \eta^2 = .34$. Both univariate group effects were significant: Digit Symbol Pairing, $F(1, 63) = 31.80, p < .001, \eta^2 = .34$; and Digit Symbol Free Recall, $F(1, 63) = 18.38, p < .01, \eta^2 = .15$. In both instances, the SIM group performed significantly worse than the NC group.

Thus with regard to the hypotheses, the simulators suppressed their performance but typically not enough to differentiate significantly from either the NC or LIT groups (see Table 7 and 8). Although the SIM group performed significantly more poorly than the NC group on a number of tasks with an attentional component (e.g., Symbol Search), no significant differences between the groups were found on the Digit Span task which has most frequently been found to be suppressed in previous research. No task differentiated the SIM group from the LIT group, although there was a non-significant

Table 8

Means and Standard Deviations for WAIS-III Subtest Raw Scores Controlling for Age and Education

Subtest	Group				
	NC n = 34	SIM n = 32	LIT n = 18	MSLIT n = 7	SUSP n = 4
	<u>M</u> (<u>SD</u>)				
D-Symb. P**	15.68 (3.29)	10.39 (4.25)	-	11.43 (4.72)	8.67 (3.06)
D-Symb. FR**	8.00 (1.07)	6.94 (1.48)	-	7.29 (1.70)	5.33 (.58)
D-Span Forw.	10.12 (2.13)	9.78 (2.56)	10.05 (1.96)	10.29 (1.70)	8.50 (2.38)
D-Span Back.	6.94 (1.87)	6.50 (2.00)	5.89 (1.70)	7.00 (2.65)	7.25 (3.59)
D-Span LF	6.71 (1.34)	6.28 (1.42)	6.63 (1.07)	6.75 (1.04)	6.00 (1.83)
D-Span LB	5.00 (1.18)	4.84 (1.22)	5.16 (3.78)	4.75 (1.58)	5.00 (2.16)
Reliable DS	9.53 (1.54)	9.50 (2.44)	9.16 (2.29)	10.14 (1.46)	8.50 (1.91)

Note: Missing data is represented by a dash (-).

D-Symb. P = Digit Symbol Pairing, D-Symb. FR = Digit Symbol Free Recall, D-Span Forw. = Digit Span Forward, D-Span Back. = Digit Span Backwards, D-Span LF = Digit Span Longest Forward, D-Span LB = Digit Span Longest Backward, Reliable DS = Reliable Digit Span.

**p < .01 for comparisons among NC, SIM, and LIT groups.

trend for the SIM group to perform even more poorly than the LIT on some tasks (e.g., Matrix Reasoning and Symbol Search). Not surprisingly, upon visual inspection the MSLIT group appeared to perform more poorly than the LIT on most tasks. The SIM

group's performance appeared more impaired than the MSLIT on Digit Span, Letter Number Sequencing, Digit Symbol, and Matrix Reasoning. Once again, the SUSP group's performance appeared to fall below the performance of all other groups.

Part V: Comparison of WAIS-III and K-BIT

Both the NC and SIM groups were asked to put forth their best effort on the K-BIT. Not surprisingly, ANOVA revealed no significant difference between the two groups on the K-BIT summary scores: Composite Score, Vocabulary, and Matrices (see Table 9).

Comparison of performance on the WAIS-III versus K-BIT is of interest, especially in the SIM group who was instructed to feign head injury for one test (WAIS-III), and put forth their best effort on the other (K-BIT). In particular, consideration of differences between these two measures allows one the opportunity to determine the source of group differences (e.g., suppression of performance versus sampling bias) on the WAIS-III summary indices previously discussed.

In the NC group, paired sample t-tests revealed significantly better performance on the WAIS-III FSIQ ($t(33) = 3.40, p < .01$) and VIQ ($t(33) = 4.33, p < .01$) but no significant difference in scores on PIQ, relative to corresponding K-BIT scores. Given the instructions that they were asked to follow, it is not surprising that similar comparisons for the SIM group revealed no significant difference in performance on the WAIS-III summary scores relative to the corresponding K-BIT scores (i.e., the difference between the WAIS-III and K-BIT approached significance for non-verbal abilities, $t(30) = -2.61, p = .01$). Thus, the SIM group's suppressed performance on the WAIS-III diluted

the expected differences that might have otherwise been observable between the two measures, as was seen in the NC group. However, comments regarding the ability of the K-BIT to estimate WAIS-III performance cannot be made based on the results from this study, as the administration of these tasks was not counterbalanced.

Table 9

Means and Standard Deviations for WAIS-III and K-BIT Summary Scores

Summary Score	Group			
	Normal Controls (n = 34)		Simulators (n = 31)	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
WAIS-III FSIQ	111.68	8.66	106.68	10.20
K-BIT Composite	108.71	7.61	107.52	7.06
WAIS-III VIQ	113.12	9.61	109.55	11.41
K-BIT Vocabulary	108.00	7.58	106.39	7.19
WAIS-III PIQ	108.41	10.45	102.10	10.56
K-BIT Matrices	107.79	8.87	107.13	7.38

Note. FSIQ = Full Scale Intelligence Quotient, VIQ = Verbal Intelligence Quotient, PIQ = Performance Intelligence Quotient.

Part VI: WMS-III Summary Scores

Means and standard deviations for the WMS-III summary scores appear in Table 10 and are displayed graphically in Figure 2. Again, one of two possible patterns of

performance was expected: overall suppression of performance or specific areas of poor performance (see part VII for latter).

A profile analysis was used to examine group (NC, SIM, LIT) differences on the WMS-III Summary scores. Once again, only the effect of parallelism was of interest⁴.

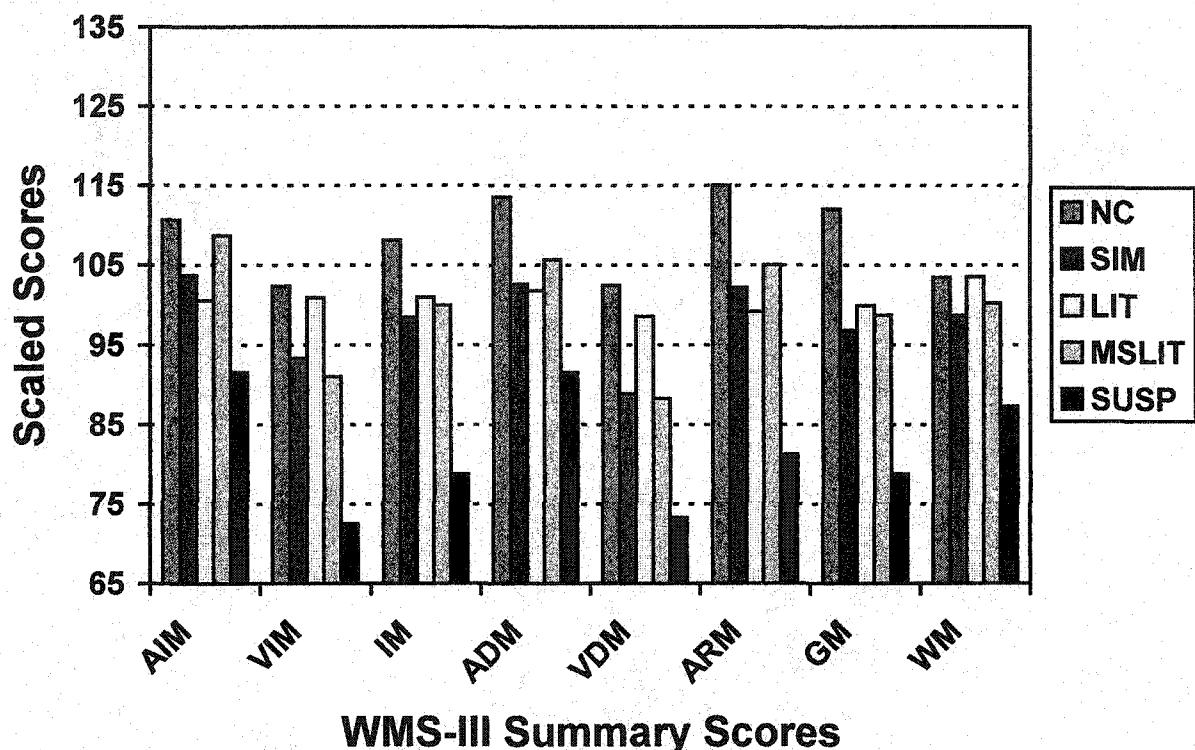


Figure 2. Group Profiles on WMS-III Summary Scores.

The test of parallelism indicated a significant interaction, $F(14, 539) = 2.70, p < .001, \eta^2 = .07$. Post-hoc comparisons (Student-Newman-Keul adjustments) were conducted to examine which groups differed on each dependent measure (significant if $\alpha < .05$).

Table 10

Means and Standard Deviations for WMS-III Summary Scores

Summary Score	Group				
	NC <u>n</u> = 30	SIM <u>n</u> = 32	LIT <u>n</u> = 19	MSLIT <u>n</u> = 8	SUSP <u>n</u> = 4
	<u>M</u> (<u>SD</u>)				
AIM*	110.67 (12.65)	103.72 (13.09)	100.58 (17.54)	108.63 (18.13)	91.50 (25.05)
VIM*	102.40 (12.99)	93.31 (15.19)	100.95 (11.64)	91.00 (11.38)	72.50 (16.05)
IM*	108.07 (11.53)	98.44 (15.13)	101.00 (13.51)	100.00 (16.14)	78.75 (24.05)
ADM*	113.50 (10.05)	102.62 (14.64)	101.84 (17.51)	105.63 (14.55)	91.50 (25.96)
VDM*	102.47 (11.94)	88.87 (14.79)	98.58 (9.71)	88.25 (10.21)	73.25 (16.70)
ARM*	115.00 (10.75)	102.19 (16.56)	99.21 (13.05)	105.00 (20.53)	81.25 (22.50)
GM*	112.00 (9.99)	96.84 (16.07)	99.95 (13.83)	98.75 (16.12)	78.75 (24.53)
WM	103.43 (10.77)	98.69 (10.61)	103.53 (14.96)	100.25 (14.61)	87.25 (14.57)

Note. AIM = Auditory Immediate Memory, VIM = Visual Immediate Memory, IM = Immediate Memory, ADM = Auditory Delay Memory, VDM = Visual Delay Memory, ARM = Auditory Recognition Memory, GM = General Memory, WM = Working Memory.

* $p < .05$. ** $p < .01$ for comparisons among NC, SIM, and LIT groups.

Both the SIM and LIT groups performed significantly more poorly than the NC group on the AIM, ADM, ARM, and GM. The SIM group also performed more poorly

than the NC group on IM, VIM, and VDM. Overall, the performance of the SIM group tended to be suppressed, and they scored significantly more poorly than the LIT group on VIM and VDM.

Casual comparisons of the MSLIT and SUSP groups, showed that the SIM group's performance at times appeared to fall below that of the MSLIT group (e.g., ADM and ARM). Although not statistically analyzed, the performance of head injured individuals suspected of biased responding (SUSP) appeared globally suppressed.

Part VII: WMS-III GMI-WMI Discrepancy Score

As reported in previous studies (e.g., Reid & Kelly, 1991), simulators sometimes show specific patterns of impairment, such as suppression of performance on attention/working memory measures relative to general memory (e.g., larger discrepancy scores). This was explored through the analysis of GMI-WMI discrepancy scores.

ANCOVA controlling for education was used to compare the three groups on GMI-WMI Discrepancy scores. The univariate comparisons of the three groups approached the specified $\alpha < .01$ criterion, $F(2, 77) = 4.71$, $p = .01$, $\eta^2 = .11$. Post-hoc contrast using Bonferroni adjustments revealed that the SIM group differed significantly from the NC group, but that they did not differ from the LIT group. Surprisingly, of all the groups (including casual visual comparison to the MSLIT and SUSP groups), only the NC group performed more poorly on the WMI than on GM.

Part VIII: WMS-III Subtest Scores

Once again, as previous researchers have observed a tendency for malingeringers to suppress overall performance, all WMS-III subtests were considered in the analyses for exploratory purposes. Different analyses were conducted on subtest scaled scores and on subtests for which only raw scores were available.

A profile analysis was used to examine group (NC, SIM, LIT) differences on the WMS-III subtest scaled scores (see Table 11 and Figure 3). Once again, only the effect of parallelism was of interest.

The test of parallelism was significant, $F(18, 693) = 2.1, p < .01, \eta^2 = .05$. For each dependent measure, post-hoc comparisons (Student-Newman-Keul) were conducted to examine all possible group combinations. All reported comparisons were significant $p < .05$.

Post-hoc contrasts revealed that the SIM group scored significantly lower than the NC on Logical Memory II, Faces II, Verbal Paired Associates II, Family Pictures I and Family Pictures II. Similarly, the LIT group performed more poorly than the NC on Logical Memory I and Logical Memory II. The SIM group performed significantly worse than the LIT group on Faces II, and there was a similar but non significant trend for the SIM group to score lower than the LIT group on Family Pictures II.

Table 11

Means and Standard Deviations for WMS-III Age-Adjusted Subtest Scores

Subtest	Group				
	NC <u>n</u> = 30	SIM <u>n</u> = 32	LIT <u>n</u> = 18	MSLIT <u>n</u> = 8	SUSP <u>n</u> = 4
	<u>M</u> (SD)	<u>M</u> (SD)	<u>M</u> (SD)	<u>M</u> (SD)	<u>M</u> (SD)
LM-I*	12.93 (2.65)	11.47 (2.85)	10.41 (3.04)	12.25 (3.65)	8.50 (4.04)
LM-II*	13.40 (2.47)	11.22 (3.07)	9.94 (2.75)	12.00 (3.70)	9.00 (4.83)
Faces-I	10.13 (2.99)	9.16 (2.75)	10.29 (3.18)	9.50 (1.69)	7.00 (2.16)
Faces-II*	10.37 (2.70)	8.16 (2.52)	11.53 (8.41)	8.25 (1.16)	7.75 (2.75)
VerbPair-I	10.80 (2.37)	9.84 (2.59)	10.06 (3.01)	10.62 (2.56)	8.75 (5.32)
VerbPair-II*	11.30 (1.51)	9.78 (2.74)	10.06 (3.40)	9.88 (1.73)	8.25 (4.11)
Family I*	10.70 (2.34)	8.78 (2.98)	9.59 (1.91)	7.63 (3.11)	4.50 (2.65)
Family II*	10.43 (2.45)	8.31 (3.02)	9.35 (2.06)	8.00 (3.63)	3.75 (2.63)
L-N Sequen.	11.10 (2.25)	10.13 (2.11)	9.71 (2.57)	10.88 (2.75)	8.25 (3.86)
Spatial Span	10.27 (2.88)	9.63 (2.25)	11.00 (3.24)	9.25 (3.88)	7.25 (2.50)

Note: LM-I = Logical Memory I, LM-II = Logical Memory II, VerbPair-I = Verbal Paired Associates I, VerbPair-II = Verbal Paired Associates II, L-N Sequen. = Letter Number Sequencing.

* $p < .05$ ** $p < .01$ for comparisons among NC, SIM, and LIT groups.

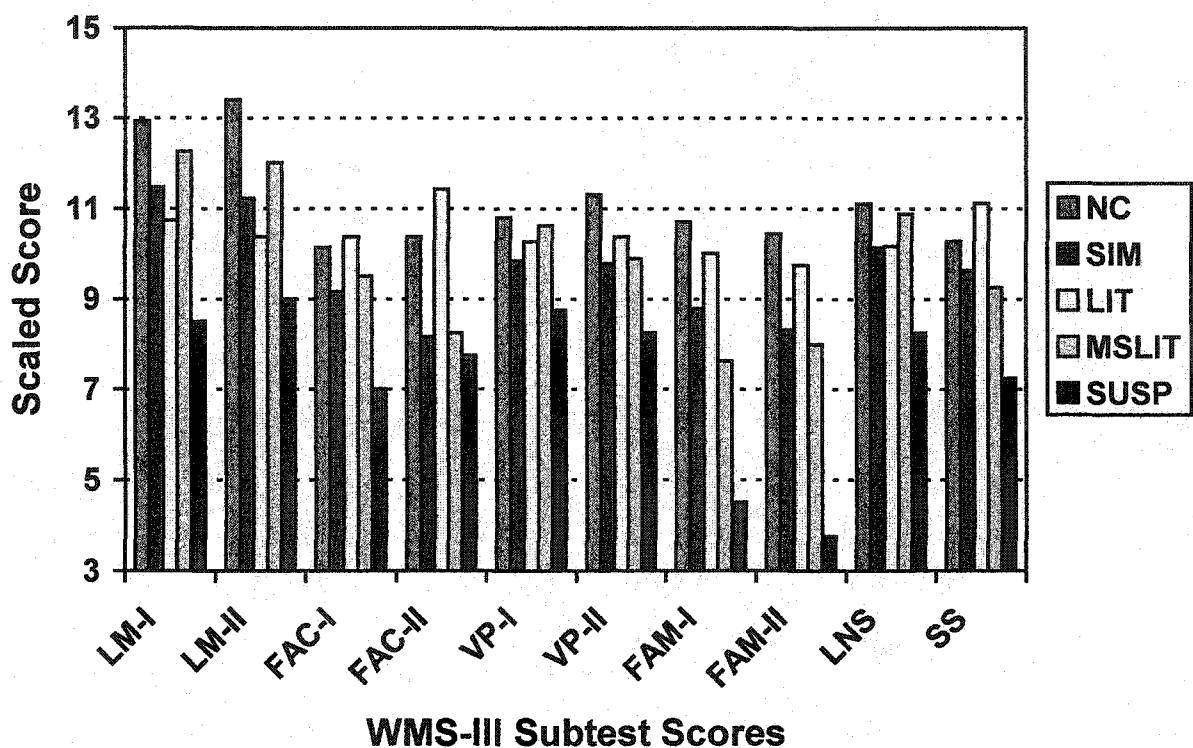


Figure 3. Group profiles on WMS-III subtests.

MANCOVA controlling for the influence of age and education, was computed on WMS-III tasks for which only raw scores were available for all three groups (see Table 12). As neither covariate yielded significant effects, the subtests were analyzed using MANOVA. At the omnibus level, the main effect of group was significant, $F(4, 158) = 4.33$, $p < .01$, $\eta^2 = .10$. Significant univariate effects for group were found for Logical Memory Recognition, $F(2, 80) = 8.60$, $p < .001$, $\eta^2 = .18$. Post-hoc contrasts revealed that both the SIM and LIT groups performed more poorly than the NC, but did not differ from each other.

Table 12

Means and Standard Deviations for WMS-III Subtest Raw Scores Controlling for Age and Education

	Group				
	NC (n = 34)	SIM (n = 32)	LIT (n = 17)	MSLIT (n = 7)	SUSP (n = 4)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Orientation*	13.47 (.71)	13.44 (.80)	-	-	-
Mental Control	12.65 (2.39)	10.94 (2.38)	-	-	-
SS Reliable	9.90 (1.70)	9.30 (1.80)	9.60 (2.00)	8.40 (2.30)	7.00 (1.40)
LM Recog**	28.20 (1.50)	26.50 (2.90)	25.70 (2.30)	25.70 (4.00)	22.80 (5.60)

Note. Missing data is represented by a dash (-).

SS Reliable = Reliable Spatial Span, LM Recog = Logical Memory Recognition

*p < .05 **p < .01 for comparisons among NC, SIM, and LIT groups.

Due to missing data for the LIT group, a separate MANOVA was computed to examine differences for the NC and SIM groups on Orientation and Mental Control. The omnibus group effect was not significant.

Overall, the performance of the SIM group was suppressed on WMS-III tasks. Although there was a trend for the SIM group to perform more poorly than the LIT group on most tasks, a significant effect was only reached on FACES II. Interestingly, looking at table 11, the SIM group also appeared to generally perform more poorly than the MSLIT group. Once again the performance of the SUSP group appeared consistently lower than that of all other groups.

Part IX: Reliable Digit Span and Reliable Spatial Span

Previous research has demonstrated that a derived score (Reliable Digit Span) based on Digit Span performance may be effective in detecting malingering (Greiffenstein et al., 1994; Strauss et al., 1999). It was expected that the SIM group would have lower Reliable Digit Span Scores and also perform more poorly on the corresponding derived score for the Spatial Span subtest (e.g., Reliable Spatial Span) than the other groups (see Table 13).

Table 13

Means and Standard Deviations for Reliable Digit Span and Reliable Spatial Span

	Group				
	NC (n = 34)	SIM (n = 32)	LIT (n = 17)	MSLIT (n = 7)	SUSP (n = 4)
	<u>M</u>	<u>M</u>	<u>M</u>	<u>M</u>	<u>M</u>
Summary Score	(SD)	(SD)	(SD)	(SD)	(SD)
Reliable Digit Span	9.53 (1.54)	9.50 (2.44)	9.24 (2.36)	10.14 (1.46)	8.50 (1.90)
Reliable Spatial Span	9.94 (1.70)	9.34 (1.77)	9.65 (1.99)	8.43 (2.30)	7.00 (1.41)

MANCOVA, controlling for the influence of age and education, was computed to compare the three groups on Reliable Digit Span and Reliable Spatial Span. As neither covariate yielded significant effects, these derived scores were analyzed using MANOVA. At the omnibus level, the main effect of group was not significant and the univariate analyses were thus not explored. It is interesting to note, however, that the SUSP group appeared to obtain the lowest Reliable Digit Span and Spatial Span scores.

Part X: Consistency of performance

Consistency of performance was examined on the Digit Span (DS) and Letter Number Sequencing (LNS) subtests appearing on both the WAIS-III and WMS-III. This resulted in a delay of over two hours between the first and second administrations of these tasks. It was hypothesized that the SIM group would have a greater difference in score from one administration to the next, than the NC group.

A 2 (Group) x 2 (Administration) repeated measures ANOVA was conducted with the NC and SIM groups, to examine whether consistency of performance across repeated administrations (WMS-III and WAIS-III) of select tasks (DS, LNS) varied as a function of group. As the primary hypothesis was concerned with the Group x Administration interaction to examine consistency, results for the main effects of group and administration are not reported. Only longest digit forward for Digit Span yielded a significant Group x Administration interaction, $F(1, 64) = 3.98, p = .05, \eta^2 = .06$, with the SIM group performing more inconsistently than the NC group. On other variables (e.g., Longest Digit Backward), the performance of the SIM group tended be more inconsistent than that of the NC group but the interactions did not reach significance. This was perhaps due to the small sample size, as revealed by the low post-hoc power estimates.

Part XI: 21-Item Test

With regard to the 21-Item Test (See Table 14), it was hypothesized that the simulators would perform more poorly than the NC group on this task which is specifically designed to detect malingering.

MANOVA was computed to examine group differences. The omnibus effect for group was significant, $F(6, 59) = 3.23, p < .01, \eta^2 = .25$. The results revealed that the

SIM group performed significantly worse than the NC group on the recognition portion of this measure (Forced Choice $F(1, 64) = 17.22, p < .001, \eta^2 = .21$). Although failing to satisfy our $\alpha = .01$ criterion, both Inconsistencies ($F(1, 64) = 4.73, p = .03, \eta^2 = .07$), and Consecutive Misses ($F(1, 64) = 5.99, p = .02, \eta^2 = .09$) approached significance with the SIM group showing a larger number of errors.

Table 14

Means and Standard Deviations for 21-Item Test Scores

	Group			
	Normal Control <u>n</u> = 34		Simulator <u>n</u> = 32	
	<u>M</u>	(<u>SD</u>)	<u>M</u>	(<u>SD</u>)
Free Recall	9.3	(3.3)	8.1	(2.7)
Forced Choice**	18.4	(1.7)	16.3	(2.5)
Inconsistencies*	0.4	(0.6)	1.0	(1.4)
Consecutive Misses*	1.4	(1.0)	2.0	(1.1)
Intrusions	0.7	(0.8)	0.6	(0.7)
Perseverations	0.3	(0.7)	0.2	(0.4)

Note. * $p < .05$, ** $p < .01$

Discussion

This study aimed to determine if individuals applying less than optimal effort could be detected through their pattern of performance on the WAIS-III, WMS-III, and a self-report measure of subjective physical and psychological complaints (BCPSI). In addition, performance on a forced-choice procedure (21-Item test) aimed specifically at assessing biased responding was assessed. Although previous research has examined simulator's patterns of performance on the older versions of the Wechsler scales (WAIS-R and WMS), few studies (e.g., Killgore & DellaPietra, 2000) to date have considered the WAIS-III and WMS-III. Therefore, limited information is available on simulator's pattern of performance on the new subtests of the Wechsler scales. In addition, this study aimed to address some of the methodological problems noted in previous research using simulation paradigms. Specifically, (1) all groups were screened for factors (e.g., alcohol problems) that could influence test results, (2) the 'head injured' group was limited to litigating individuals with reported head injury resulting from external factors, (3) individuals suspected of feigning deficits (e.g. SUSP group) were excluded from the litigating group (4) more detailed simulation instructions were used, (5) simulators were provided with instructions in advance thereby providing them with the opportunity to think about the sequelae of head injury and malingering techniques, and (6) a compliance check was completed.

The results from this study revealed that the SIM group endorsed more subjective concerns than the NC group on a self-report measure of physical and psychological complaints (BCPSI) often reported following head injury. Specifically, the two groups differed in their experience of headaches, dizzy spells, fatigue, visual problems, sleep

disturbances, memory problems, concentration problems, and irritability. The simulators' performance on this measure suggests an understanding of head injury and compliance with test instructions. Unfortunately, comparison with the litigants was not possible.

With regard to cognitive functioning, comparison of performance on the WAIS-III to performance on the KBIT revealed that the NC group performed better on the WAIS-III FSIQ and VIQ relative to corresponding K-BIT scores. Similar comparisons for the SIM group revealed no significant difference in performance on the WAIS-III summary scores relative to the corresponding K-BIT scores, likely reflecting their adherence to the malingering instructions during completion of the WAIS-III. Thus, it appeared that the SIM group's suppressed performance on the WAIS-III diluted the expected differences that might have otherwise been observable between the two measures, as was seen in the NC group.

Although results from the SIM group were often suppressed on the WAIS-III, their performance did not typically differ significantly from the performance of either the NC or LIT groups. Specifically, on the summary scores of the WAIS-III, the SIM group performed more poorly than the NC only on PIQ and PSI, and differed from the LIT group only on VIQ and VCI. Unlike previous research (e.g., Binder & Willis, 1991; Mittenberg et al., 1993; Trueblood, 1994), the SIM group did not suppress their performance to such an extent that they actually performed significantly worse than the LIT group.

With regard to specific patterns of impairment, no significant differences were found between the groups on the VIQ-PIQ Discrepancy Scores. Interestingly, however, the LIT group was unexpectedly found to perform better on PIQ than VIQ. On the

Vocabulary-Digit Span Discrepancy Score, the NC group was surprisingly found to have the largest discrepancy score.

At the subtest level, the results once again revealed that the SIM group suppressed their performance but typically not enough to differ significantly from the NC or LIT groups. Specifically, the SIM group only differed from the NC on Digit Symbol (Total, Pairing, and Free Recall) and Symbol Search. No task differentiated the SIM group from the LIT group.

The SIM group's pattern of performance on memory tasks (WMS-III) was more in keeping with the overall suppressed performance reported in previous research. Specifically, on the WMS-III summary scores, the SIM group performed significantly more poorly than the NC group on AIM, ADM, ARM, GM, IM, VIM, and VDM. In addition, like previous reports of excessively poor performance (e.g., Klimczak et al., 1997), the SIM group in this study performed more poorly than the LIT group on VIM and VDM.

An unusual pattern of performance was observed on GMI-WMI discrepancy scores. Previous studies have revealed that simulators tend to suppress their performance on attention/working memory measures relative to general memory (e.g., Reid & Kelly, 1991). This pattern was observed in the NC group not the SIM and LIT groups.

At the subtest level, the SIM group performed more poorly than the NC group on Logical Memory II, Faces II, Verbal Paired Associates II, as well as Family Pictures I and II. Once again, the SIM group suppressed their performance excessively and performed significantly more poorly than the LIT on Faces II.

Although previous research has demonstrated the potential utility of a derived score (Reliable Digit Span) in detecting simulators, similar results were not obtained in this study. In addition, a derived score based on Spatial Span was also not effective in detecting biased responding.

With regard to the usefulness of consistency of performance in the detection of biased responding, results from this study only revealed a group by interaction effect for Longest Digit Forward, with the SIM group performing more inconsistently than the NC group.

Although the MSLIT and SUSP groups were not included in the analyses due to small sample sizes, the data from these groups proved to be interesting for casual comparisons. Not surprisingly, the MSLIT group was found to typically perform more poorly than the LIT group. As reported in previous research (e.g., Trueblood, 1994), the SIM group's suppressed performance sometimes resembled that of the MSLIT group and was even lower than the MSLIT on occasion.

The performance of the SUSP group was of particular interest. As described in the methods section, this group was comprised of four litigating individuals reporting a history of head injury (three mild, one moderate to severe head injury) who, for the purpose of this study, were classified as having put forth "suspect" effort based on their performance on at least one test specifically assessing biased responding (other symptom validity tests passed). In this study, the performance of the SUSP group was consistently found to be lower than all of the other groups, including that of litigants with a history of more severe head injury (MSLIT).

Conclusion

In keeping with previous research (e.g., Klimczak et al., 1997), results from this study revealed that when individuals feign symptoms of head injury, they tend to suppress their performance, especially on tasks of learning and memory. However, unlike previous reports (e.g., Mittenberg et al., 1993), simulators did not typically suppress their performance to such an extent that they scored more poorly than individuals reporting a history of head injury. This perhaps reflects the use of more detailed instructions, which specified that simulators' efforts to fake symptoms should not be obvious or easily detectable by the experimenter. Alternatively, the results may be due, at least in part, to preparation time provided to the simulators. Previous research designs have not granted participants the opportunity to prepare for the testing session. In this study, the simulators received the instructions at least 24 hours prior to the evaluation, thereby attempting to more closely resemble real-life situations which allow individuals time to prepare for the assessment. It was hypothesized that providing the SIM group with preparation time would enable them to be less obvious in their feigning of deficits and consequently harder to detect; that they would be more selective or focused in their malingering. The lack of differences between the SIM group and LIT group in this study, suggests that under certain conditions, simulators may be able to effectively feign persisting complaints sometimes associated with mild head injury.

Specific markers of biased responding identified in previous research (e.g., Vod DS, reliable DS, consistency of performance) were not useful in detecting malingered performance in this study. These results are in keeping with more recent findings that

Voc-DS discrepancy may be insensitive and non-specific as an indicator of biased responding (Strauss et al., 1999). Similarly, Strauss and colleagues (1999) also found that simulators may be able to replicate their performance when the tasks are brief. When tasks are longer or when the administrations are separated by longer delays (e.g., weeks or months), consistent performance may be less likely to occur in simulators. In addition, qualitative versus quantitative differences from one administration to the next, may be prove useful in differentiating biased responders from truly head injured individuals. These differences were not assessed in the current study.

A number of factors must be considered in interpreting the results of this study and in considering future research on biased responding. One consideration is that both the NC and SIM groups were derived from a homogenous sample of educated individuals (e.g., university students). As such, they may not be representative of individuals who typically present for evaluation following head injury. Although there may be some concern that highly educated individuals may feign symptoms of head injury differently than less educated individuals, a study conducted by Iverson and Frazen (1994) revealed that undergraduate simulators had patterns similar to inmates asked to malinger. Also see recent work by Strauss and colleagues (in press), who found that experience of participants had little impact on performance.

Another consideration is that the NC group in this study obtained some unexpected results (e.g., Voc- DS score). Klimczak and colleagues (1997) point out that there may be reason to question the motivation of subjects recruited through undergraduate subject pools. In their study, normal controls performed below normative levels across all scores, thereby raising questions regarding the level of effort put forth by

these subjects. However, although unexpected scores were obtained from the NC group in this study, their overall performance did fall within the normal range suggesting adequate motivation. Thus the results obtained by the NC group may reflect sampling bias.

In addition, when interpreting the results of this study, the extent of similarities between simulators and true malingering must be considered. A number of factors inherent in real life situations of malingering cannot be easily replicated in experimental designs. Specifically, the occurrence of a traumatic event, contact with numerous sources of information, the personal and emotional investment, negative incentives, and the motivation to malinger (e.g., large financial gains).

Lastly, the LIT group in this study did perform more poorly than the NC group, despite the fact that 90 percent of them were tested over three months after their injury, when most symptoms of mild head injury should have typically resolved. This is not surprising given that all litigants were symptomatic when evaluated. However, it is important to note that a number of factors may have contributed to the results. Although actual brain injury may have influenced the performance of some individuals in this group, the results could have been affected by emotional factors (e.g., depression or anxiety), medical conditions (e.g., chronic pain), premorbid personality factors, or the effects of prescribed medication (e.g., see Klonoff & Lamb, 1998; Suhr et al., 1997). The possible individual or additive effects of these factors were not addressed in this study.

With regard to future directions in research, most studies to date have been done using positive incentives (Rogers & Cruise, 1998). The impact of negative incentives (e.g., avoiding humiliation or embarrassment) on biased responding warrants further

investigation. In addition, more research on preparation techniques and malingering strategies is also needed to further our understanding of biased responding. To date, only two studies have evaluated test-taking strategies (Cochrane, Baker, & Meudell, 1998; Iverson et al., 1995). Less than four percent of the respondents in the study conducted by Iverson and colleagues (1995) described a specific method of preparation, and over 15 percent of the participants described faking total amnesia.

Future research is also needed to address the impact of personality and/or cognitive factors on biased responding, and to explore gender differences in approaches taken to malingering. With regard to gender differences, although gender did not appear to influence patterns of biased responding in this study, Klimczak and colleagues (1997) found that women displayed more variability between measures than did men who typically had flatter profiles.

In summary, due to the increasing role of neuropsychologists in the medical legal context, the challenges posed by minor head injury, and the limited resources available for head-injured individuals, proper diagnosis of head injury and assessment of biased responding remains paramount. Although research in this area has focused on the use of specific tests of biased responding or pattern analysis on existing standardized tests, this study as well as others (see Slick et al., 1999 for review) suggests that either approach, in isolation, is not sufficient. Rather, research findings are increasingly emphasizing the importance of incorporating multiple detection measures to assure accurate assessment and diagnosis of head injury. In considering biased responding as an explanation for test results, clinicians should consider (1) performance on specific tests of malingering, (2) intra and inter test performance in order to evaluate inconsistencies or nonsensical

patterns in the results, (3) self-report of symptoms, (4) information provided by collaborative informants, and (5) behaviours outside of the testing session.

In addition, it is important to remember that the presence of biased responding does not negate the possible presence of actual deficits. Furthermore, poor effort on evaluations may be due to factors other than malingering. Specifically, the results may be due to fatigue, defensive or hostile test taking behaviour, legitimate neurological impairment, psychiatric disturbance, chronic pain, substance abuse or medications. Therefore, when the obtained results suggest that less than optimal effort was put forth by the client, the underlying cause(s) for these results must be carefully considered.

References

- American Psychiatric Association. (1994). Diagnostic and Statistical Manual for Mental Disorders (4th ed.). Washington, D.C.: Author.
- Alliger, G.M., Lilienfeld, S.O., & Mitchell, K.E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. Psychological Science, 7, 32-39.
- Aubrey, J.C., Dobbs, A.R., & Rue, B.G. (1989). Laypersons' knowledge about the sequelae of minor head injury and whiplash. Journal of Neurology, Neurosurgery, and Psychiatry, 52, 842-846.
- Baker, G.A., Hanley, J.R., Jackson, H.F., Kimmance, S., & Slade, P. (1993). Detecting the faking of amnesia: Performance differences between simulators and patients with memory impairment. Journal of Clinical and Experimental Neuropsychology, 15, 668-684.
- Bernard, L.C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. Journal of Clinical and Experimental Neuropsychology, 12, 715-728.
- Bernard, L.C. (1991). The detection of faked deficits on the Rey Auditory Verbal Learning Test: the effect of serial position. Archives of Clinical Neuropsychology, 6, 81-88.
- Bernard, L.C., McGrath, M.J., & Houston, W. (1993). Discriminating between simulated malingering and closed head injury on the Wechsler Memory Scale-Revised. Archives of Clinical Neuropsychology, 8, 539-551.
- Bianchini, K.J., Mathias, C.W., & Greve, K.W. (2001). Symptom validity testing: a critical review. The Clinical Neuropsychologist, 15, 19-45.
- Binder, L.M. (1986). Persisting symptoms after mild head injury: a review of the postconcussive syndrome. Journal of Clinical and Experimental Neuropsychology, 8, 323-346.
- Binder, L.M. (1990). Malingering following minor head trauma. The Clinical Neuropsychologist, 4, 25-36.
- Binder, L.M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. Journal of Clinical and Experimental Neuropsychology, 15, 170-182.

- Binder, L.M. (1997). A review of mild head trauma. Part II: Clinical implications. Journal of Clinical and Experimental Neuropsychology, 19, 432-457.
- Binder, L.M., & Pankratz, L. (1987). Neuropsychological evidence of a factitious memory complaint. Journal of Clinical and Experimental Neuropsychology, 9, 167-171.
- Binder, L.M., & Rohling, M.L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head-injury. American Journal of Psychiatry, 153, 7-10.
- Binder, L.M., Rohling, M.L., & Larrabee, G.J. (1997). A review of mild head trauma. Part I: meta-analytic review of neuropsychological studies. Journal of Clinical and Experimental Neuropsychology, 19, 421-431.
- Binder, L.M., Villanueva, M.R., Howieson, D., & Moore, R.T. (1993). The Rey AVLT Recognition Memory Task measures motivational impairment after mild head trauma. Archives of Clinical Neuropsychology, 8, 137-147.
- Binder, L.M., & Willis, S.C. (1991). Assessment of motivation after financially compensable minor head trauma. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3, 175-181.
- Boyer, C.L. (1991). Wechsler Memory Scale-Revised: Performance following severe closed head injury. The Clinical Neuropsychologist, 5, 262.
- Brandt, J. (1988). Malingering amnesia. In R. Rogers (Ed.), Clinical assessment of malingering and deception (pp. 65-83). New York: Guilford Press.
- Brandt, J., Rubinsky, E., & Lassen, G. (1985). Uncovering malingering amnesia. Annals of the New York Academy of Sciences, 444, 502-503.
- Cochrane, H.J., Baker, G.A., & Meudell, P.R. (1998). Simulating a memory impairment: can amnesics implicitly outperform simulators? British Journal of Clinical Psychology, 37, 31-48.
- Crossen, J.R., & Wiens, A.N. (1988). Residual neuropsychological deficits following head injury on the Wechsler Memory Scale-Revised. The Clinical Neuropsychologist, 2, 393-399.
- Cullum, C., Heaton, R., & Grant, I. (1991). Psychogenic factors influencing neuropsychological performance: somatoform disorders, factitious disorders, and malingering. In H.O. Doerr & A.S. Carling (Eds.), Forensic Neuropsychology: Legal and Scientific Bases (pp 141-171). New York: Guilford Press.

DiCarlo, M.A., Gfeller, J.D., & Oliveri, M.V. (2000). Effects of coaching on detecting feigned cognitive impairment with the category test. Archives of Clinical Neuropsychology, 15, 399-413.

Dikmen, S., McLean, A., & Temkin, N. (1986). Neuropsychological and psychosocial consequences of minor head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 49, 1227-1232.

Ekman, P., & O'Sullivan, M. (1993). Who is misleading whom? A reply to Nickerson and Hammond. American Psychologist, 48, 989-990.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: a general power analysis program. Behaviour Research Methods, Instruments, and Computers, 28, 1-11.

Faust, D. & Ackley, M.A. (1998). Did you think it was going to be easy? Some methodological suggestions for the investigation and development of malingering detection techniques. In C. R. Reynolds (Ed.). Detection of Malingering During Head Injury Litigation. 1-54. New York: Plenum Press.

Faust, D., Hart, K., Guilmette, T.J., & Arkes, H.R. (1988). Neuropsychologists' capacity to detect adolescent malingering. Professional Psychology: Research and Practice, 14, 508-545.

Faust, D., & Guilmette, T.J. (1990). To say it's not so doesn't prove that it isn't: Research on the detection of malingering. Reply to Bigler. Journal of Consulting and Clinical Psychology, 58, 248-250.

Fox, D.D., Lees-Haley, P.R., Earnest, K., & Dolezal-Wood, S. (1995). Base rates of post-concussive symptoms in HMO patients and controls. Neuropsychology, 9, 606-611.

Frazen, M.D., & Iverson, G.L. (1997). The detection of biased responding in neuropsychological assessment. In A. MacNeill-Horton, D. Wedding, J. Webster (Eds.). The Neuropsychology Handbook. Springer Publishing Company Check

Frazen, M.D., Iverson, G.I., & McCracken, L.M. (1990). The detection of malingering in neuropsychological assessment. Neuropsychology Review, 1, 247-279.

Fredrick, R.I., & Foster, H.G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. Psychological Assessment, 3, 596-602.

Fredrick, R.I., Sarfaty, S.D., Johnston, D., & Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. Neuropsychology, 8, 118-125.

- Gouvier, W.M.D., Hayes, J.S., & Smiroldo, B.B. (1998). The significance of base rates, test sensitivity, test specificity, and subjects' knowledge of symptoms in assessing TBI sequelae and malingering. In C. Reynolds (Ed.), Detection of Malingering During Head Injury Litigation. (pp. 55-79). New York: Plenum Press.
- Greiffenstein, M.F., Baker, J.W., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. Psychological assessment, 6, 218-224.
- Greiffenstein, M.F., Baker, J.W., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and postconcussion syndrome. Journal of International Neuropsychological Society, 2, 477-485.
- Gronwall, D. (1991). Minor head injury. Neuropsychology, 5, 253-265.
- Grote, C.L., Kooker, E.K., Garron, D.C., Nyenhuis, D.L., Smith, C.A., & Mattingly, M.L. (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symtpom Validity Test: cross-validation and extension of a strandardization study. Journal of Clinical and Experimental Neuropsychology, 22, 709-719.
- Guilmette, T.J., Faust, D., Hart, K., & Arkes, H.R. (1990). A national survey of psychologists who offer neuropsychological services. Archives of Clinical Neuropsychology, 5, 373-392.
- Gutierrez, J.M., & Gur, R.C. (1998). Detection of Malingering Using Forced-Choice Techniques. In C. R. Reynolds (Ed.). Detection of Malingering During Head Injury Litigation. 81-104. New York: Plenum Press.
- Haines, M.E., & Norris, M.P. (1995). Detecting the malingering of cognitive deficits: An update. Neuropsychology Review, 5, 125-148.
- Hayes, J.S., Martin, R., & Gouvier, W.D. (1995). Influence of prior knowledge and experience on the ability to feign mild head injury symptoms in head injured and non-head-injured college students. Applied Neuropsychology, 2, 63-66.
- Hayward, L., Hall, W., Hunt, M., & Zubrick, S.R. (1987). Can localized brain impairment be simulated on neuropsychological test profiles? Australian and New Zealand Journal of Psychiatry, 21, 87-93.
- Heaton, R.K., Smith, H.H., Lehman, R.A.W., & Vogt, A.T. (1978). Prospects for faking believable deficits on neuropsychological testing. Journal of Consulting and Clinical Psychology, 46, 892-900.
- Iverson, G.L., & Frazen, M.D. (1996). Using multiple objective memory procedures to detect simulated malingering. Journal of Clinical and Experimental Neuropsychology, 18, 38-51.

- Iverson, G.L., Frazen, M.D., & McCracken, L.M. (1991). Evaluation of a standardized instrument for the detection of malingered memory deficits. Law and Human Behavior, 15, 667-676.
- Iverson, G.L., Frazen, M.D., & McCracken, L.M. (1994). Application of a forced-choice memory procedure designed to detect experimental malingering. Archives of Clinical Neuropsychology, 9, 437-450.
- Iverson, G.L., Slick, D.J., & Frazen, M.D. (2000). Evaluation of WMS-R malingering index in a non-litigating clinical sample. Journal of Clinical and Experimental Neuropsychology, 22, 191-197.
- Johnson, J.L., Bellah, C.G., Dodge, T., Kelley, W., & Livingston, M.M. (1998). Effect of warning on feigned malingering on the WAIS-R in college samples. Perceptual and Motor Skills, 87, 152-154.
- Johnson, J.L., & Lesniak-Karpia, K. (1997). The effect of warning on malingering of memory and motor tasks in college samples. Archives of Clinical Neuropsychology, 12, 231-238.
- Kaufman, A.S., & Kaufman, N.L. (1990). Kaufman Brief Intelligence Test Manual. Circle Pines, MN: American Guidance Service.
- Kay, T., Newman, B., Cavallo, M., Ezrachi, O., & Resnick, M. (1992). Toward a neuropsychological model of functional disability after mild traumatic brain injury. Neuropsychology, 4, 371-384.
- Killgore, W.D.S., & DellaPietra, L. (2000). Using the WMS-III to detect malingering: empirical validation of the rarely missed index (RMI). Journal of Clinical and Experimental Neuropsychology, 22, 761-771.
- Klimczak, N.J., Donovick, P.J., & Burright, R. (1997). The malingering of multiple sclerosis and mild traumatic brain injury. Brain Injury, 11, 343-352.
- Klonoff, P.S., & Lamb, D.G. (1998). Mild head injury, significant impairment on neuropsychological test scores, and psychiatric disability. The Clinical Neuropsychologist, 12, 31-42.
- Larrabee, G.J. (1991). Cautions in the use of neuropsychological evaluation in legal settings: Neuropsychological tests can be failed for reasons other than brain damage. Neuropsychology, 4, 239-249.
- Levin, H.S. (1989). Memory deficit after closed-head injury. Journal of Clinical and Experimental Neuropsychology, 12, 129-153.

Levin, H.S. & Goldstein, F.C. (1986). Organization of verbal memory after severe closed-head injury. Journal of Clinical and Experimental Neuropsychology, 8, 643-656.

Lezak, M.D. (1995) Neuropsychological Assessment. New York: Oxford University Press.

LoPiccolo, C.J., Goodkin, K., & Baldewicz, T.T. (1999). Current issues in the diagnosis and management of malingering. Annals of Medicine, 31, 166-174.

Mandleberg, I.A., & Brooks, D.N. (1975). Cognitive recovery after severe head injury: 1. Serial testing on the Wechsler Adult Intelligence Scale. Journal of Neurology, Neurosurgery, and Psychiatry, 38, 1121-1126.

Martin, R.C., Bolter, J.F., Todd, M.E., Gouvier, W.D., & Nicolls, R. (1993). Effects of sophistication and motivation on the detection of malingered memory performance using a computerized forced-choice task. Journal of Clinical and Experimental Neuropsychology, 15, 867-880.

Martin, N.J., & Franzen, M.D. (1992). The malingering of memory deficits among sophisticated subjects. The Clinical Neuropsychologist, 6, 329.

Mendelson, G. (1995). Compensation neurosis revisited: outcome studies of the effects of litigation. Journal of psychosomatic Research, 39, 695-706.

Miller, L. (1990). Litigating the head trauma case: Issues and answers for attorneys and their clients. Cognitive Rehabilitation 8, 8-12.

Miller, W.G., & Miller, S.E. (1992). Malingering and neuropsychological assessment. Physical Medicine and Rehabilitation: State of the Art Reviews, 6, 547-563.

Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale-Revised. Psychological Assessment, 5, 34-40.

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.

Neis, 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.

Osimani, A., Alon, A., Berger, A., & Abarbanel, J.M. (1997). The use of the Stroop phenomenon as a diagnostic tool for malingering. Journal of Neurology, Neurosurgery, and Psychiatry, 62, 617-621.

- Overholser, J.C. (1990). Differential diagnosis of malingering and factitious disorder with physical symptoms. Behavioral Sciences and the Law, 8, 55-65.
- Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. Perceptual and Motor Skills, 57, 367-372.
- Price, J.R., & Stevens, K.B. (1997). Psycholegal implications of malingered head trauma. Applied Neuropsychology, 4, 75-83.
- Psychological Corp. (1997). The Wechsler Adult Intelligence Scale- III. San Antonio, TX: Harcourt Brace.
- Psychological Corp. (1997). The Wechsler Memory Scale- III. San Antonio, TX: Harcourt Brace.
- Rawlings, D.B., & Crewe, N.M. (1992). Test-retest practice effects and test score changes of the WAIS-R in recovering traumatically brain-injured survivors. The Clinical Neuropsychologist, 6, 415-430.
- Reid, D.B., & Kelly, M.P. (1991). A study of the Wechsler Memory Scale-Revised in closed head injury. Journal of Clinical and Experimental Neuropsychology, 13, 20.
- Reitan, R.M., & Wolfson, D. (1985). Neuroanatomy and neuropathology: a clinical guide for neuropsychologists, Tuscon, AZ: Neuropsychology Press.
- Reitan, R.M., & Wolfson, D. (1995). Consistency of response on retesting among head-injured subjects in litigation versus head-injured subjects not in litigation. Applied Neuropsychology, 2, 67-71.
- Reitan, R.M., & Wolfson, D. (1996). Wetter, M.W., & Deitsch, S.E. (1996). Faking specific disorders and temporal response consistency on the MMPI-2. Psychological Assessment, 8, 39-47.
- Reitan, R.M., & Wolfson, D. (1996). The question of validity of neuropsychological test scores among head-injured litigants: development of a dissimulation index. Archives of Clinical Neuropsychology, 11, 573-580.
- Reitan, R.M., & Wolfson, D. (1997). Consistency of neuropsychological test scores of head-injured subjects involved in litigation compared with head-injured subjects not involved in litigation: Development of the Retest Consistency Index. The Clinical Psychologist, 11, 69-76.
- Resnick, P.J. (1988). Malingering of posttraumatic disorders. In R. Rogers (Ed.), Clinical Assessment of Malingering and Deception (pp. 84-107). New York: Guilford.

- Reynolds, C.R. (1998). Common sense, clinicians, and actuarialism. In C.R. Reynolds (Ed.), Detection of malingering during head injury litigation. New York: Plenum.
- Rogers, R. (Ed.). (1997). Clinical assessment of malingering and deception (2nd ed.). New York: Guilford Press.
- Rogers, R. & Cruise, K.R. (1998). Assessment of malingering with simulation designs: threats to external validity. Law and Human Behaviour, 22, 273-285.
- Rogers, R., Harrell, E.H., & Liff, C.D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. Clinical Psychology Review, 13, 255-274.
- Rose, F.E., Hall, S., Szalda-Petree, A.D., & Bach, P.J. (1998). A comparison of four tests of malingering and the effects of coaching. Archives of Clinical Neuropsychology, 13, 349-363.
- Ruff, R.M., & Richardson, A.M. (1999). Mild traumatic brain injury. In J.J. Sweet (Ed.). Forensic Neuropsychology: Fundamentals and Practice. 315-338. New York: Swets & Zeitlinger.
- Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, H.L. (1998). Cognitive complaints in patients after whiplash injury: the impact of malingering. Journal of Neurology, Neurosurgery, and Psychiatry, 64, 339-343.
- Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingering, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's forced-choice memory test. Journal of Clinical and Experimental Neuropsychology, 16, 472-481.
- Slick, D., Hopp, G., Strauss, E., & Thompson, G. (1997). Victoria Symptom Validity Test. Odessa, FL: Psychological Assessment Resources.
- 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.
- Spreen, O. & Strauss, E. (1998). A compendium of neuropsychological tests: Administration, norms, and commentary, (2nd ed.), New York: Oxford University Press.
- Strauss, E., Hultsch, D.F., Hunter, M., Slick, D.J., Patry, B., & Levy-Bencheton, J. (2000). Using intraindividual variability to detect malingering in cognitive performance. The Clinical Neuropsychologist, 14, 420-432.

Suhr, J.A. & Boyer, D. (1999). Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. Journal of Clinical and Experimental Neuropsychology, 21, 701-708.

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

Suhr, J., Tranel, D., Wefel, J., & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. Journal of Clinical and Experimental Neuropsychology, 19, 500-514.

The Psychological Corporation. (1997). WAIS-III and WMS-III Technical Manual. San Antonia TX: Harcourt Brace & Company.

Thompson, L.L., & Cullum, C.M. (1991). Pattern of performance on neuropsychological tests in relation to effort in mild head injury patients. Archives of Clinical Neuropsychology, 6, 231.

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

Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered and other invalid WAIS-R and clinical memory data. Journal of Clinical and Experimental Neuropsychology, 16, 597-607.

Trueblood, W. & Binder, L.M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingers. Archives of Clinical Neuropsychology, 12, 13-27.

Trueblood, W. & Frank, R.J. (1994). Detecting malingering in a minor closed head injury case: Objective neuropsychological methods. Trial Diplomacy Journal, 17, 65-72.

Trueblood, W., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. Journal of Clinical and Experimental Neuropsychology, 15, 578-590.

Wasyliw, O.E., & Cavanaugh, J.L. (1989). Simulation of brain damage: Assessment and decision rules. Bulletin of the American Academy of Psychiatry and Law, 17, 373-386.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale: Third Edition. San Antonio, TX: The Psychological Corporation.

- Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. Archives of Clinical Neuropsychology, 4, 233-265.
- Wiggins, E.C. & Brandt, J. (1988). The detection of simulated amnesia. Law and Human Behaviour, 12, 57-78.
- Youngjohn, J.R. (1995). Confirmed attorney coaching prior to neuropsychological evaluation. Assessment, 2, 279-283.
- Youngjohn, J.R., Lees-Haley, P.R., & Binder, L.M. (1999). Warning malingering produces more sophisticated malingeringers. Archives of Clinical Neuropsychology, 14, 511-515.
- Zappala, G., & Trexler, L.E. (1992). Quantitative and qualitative aspects of memory performance after minor head injury. Archives of Clinical Neuropsychology, 7, 145-154.

Footnotes

¹ A 2 (Group) x 2 (Gender) multivariate analysis of variance was computed to determine the possible effect of gender on the pattern of performance on BCPSI items. The Group x Gender interaction was not significant.

² A 2 (Group) x 2 (Gender) multivariate analysis of variance was computed to determine the possible effect of gender on the pattern of performance on the WAIS-III summary scores. The Group x Gender interaction was not significant.

³ As the LIT group did not complete either the Picture Arrangement or Comprehension subtests of the WAIS-III, a separate MANOVA was computed to examine differences between the SIM and NC groups. The omnibus group effect for this analysis was not significant.

⁴ A 2 (Group) x 2 (Gender) multivariate analysis of variance was computed to determine the possible effect of gender on the pattern of performance on the WMS-III summary scores. The Group x Gender interaction was not significant.