Detecting Malingering Using Reaction Time Tests:

Development of a New Paradigm?

By

Janna Kost

A thesis submitted to
The Faculty of Graduate Studies and Research
In partial fulfillment of the requirements for the degree of
Masters of Arts on Psychology

Department of Psychology

Carleton University
Ottawa, Ontario
November 2003

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitions et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-89847-4

Our file Notre référence

ISBN: 0-612-89847-4

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this dissertation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de ce manuscrit.

While these forms may be included in the document page count, their removal does not represent any loss of content from the dissertation.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

With increasing public familiarity of current methods of detecting malingering new, improved, and less well-known ways of detection are required. One purpose of this study was to examine the Computer Test of Information Processing (CTIP), a pure reaction time task, as a new method to detect malingering. The CTIP is made of three reaction time tasks that vary in complexity. Four groups (Control, Malingering, Mild Traumatic Brain Injury, and Moderate Traumatic Brain Injury) were used in this part of the study. Results revealed that the malingering group 'overestimated' the response time performance of someone with a head injury on the CTIP tests. For all levels of complexity the Malingering Group obtained longer reaction time scores and increased variability than did all other groups. The other purpose of the study was to compare three conventional detection methods and to examine whether latency scores could be used to detect malingering. The three conventional tests used in the study were the Test of Memory Malingering, the 21-Word Test, and the Victoria Symptom Validity Test. The TBI groups were not used in this part of the study. Examination of correct number of responses showed a lower sensitivity rating than previous studies have found for all three conventional tests. Reaction time seems to be an effective, subtle way to detect individuals attempting to malingering a head injury in that the Malingering Group showed longer reaction time scores than the Control Group. This suggests that the use of reaction time scores may be a viable way of detecting malingering.

Acknowledgements

It has been an honour to have worked with Dr. Tom Tombaugh, to whom I would like to extend a sincere thank you. His guidance and enthusiasm for research has taught me a great deal. Additionally, I would like to thank my father, Corrie Kost, for all his support and encouragement leading up to and through out my MA program. Finally, I would like to thank Grant for his outstanding patience during the past few years.

Table of Contents

Abstract	III
Acknowledgment	IV
Table of Contents	V
List of Tables	VI
List of Figures	VII
Introduction	1
Why is the Study of Malingering Important?	4
History	6
Early Malingering Tests	12
Symptom Validity Tests (SVT)	15
Methodological Variations of the SVT	18
Computerized techniques aimed at malingering detection	20
Some critical issues in the study of malingering	21
Why Further Test Development is Necessary	25
Reaction Time Measures of Malingering	27
Reaction Time Tests and Malingering	33
Summary	34
Method	36
Results	41
Computer Test of Information Processing	41
Conventional Malingering Tests	53
Discussion	62
Computer Test of Information Processing	63
Conventional Malingering Tests	73
Strategies	87
Future Research	87
References	89
Appendix	98

List of Tables

Table 1:	Means and Confidence Intervals for each CTIP test for all groups	43
Table 2:	Semantic categorical responses (ms) on the CTIP	44
Table 3:	Multiple comparison results for percent ratio change on CTIP semantic task	46
Table 4:	Sensitivity/Specificity ratings for CTIP latency scores	47
Table 5:	Semi-interquartile range for CTIP tests	49
Table 6:	Ratio of semi-interquartile to median scores for each group on the CTIP tests	50
Table 7:	Number of correct responses and latencies (ms) for the Word Memory Test	56
Table 8:	Specificity/sensitivity ratings for conventional malingering tests	61

List of Figures

Figure 1:	Reaction time (ms) on the CTIP for malingerers, controls, mild TBI and moderate TBI groups	42
Figure 2:	Percent change for same and different responses on semantic search	45
Figure 3:	Percentile Rank for simple CTIP	52
Figure 4:	Percentile Rank for choice CTIP	52
Figure 5:	Percentile Rank for semantic CTIP	52
Figure 6:	Number of correct responses for each trial on the TOMM for the Malingering and Control groups	53
Figure 7:	Response latencies for all responses on the TOMM for the Malingering and Control groups	54
Figure 8:	Response latencies for correct and incorrect responses on the TOMM for the Malingering and Control groups	54
Figure 9:	Number of correct responses on each trial for the Malingering and Control Groups	57
Figure 10:	Number of correct response for easy and hard number on each trial for the Malingering and Control groups	57
Figure 11:	Overall latency scores for each trial on the Victoria	58
Figure 12:	Easy and Hard latencies for malingering and control group on the Victoria	59
Figure 13:	Correct/Incorrect Latencies for malingering and control Group For each trial on the Victoria	59

INTRODUCTION

The purpose of the current study is two-fold. The first purpose is to show how three different reaction time tests can be used to detect malingering. The second intent is to compare the performance on these reaction time tasks to that of conventional malingering tests (Test of Memory Malingering, TOMM; Tombaugh, 1996; 21-word item test, Iverson, Franzen, & McCracken, 1991; Victoria Symptom Validity Test, VSVT; Slick, Hoop, Strauss, Hunter & Pinch, 1994).

Malingering refers to the “intentional production of false or grossly exaggerated physical or psychological symptoms that are motivated by external incentives such as financial gain, reduced workload, and escape from criminal conviction” (American Psychiatric Association, 1994, p. 683). Although people are often referred to as being “malingerers” there is very little evidence to suggest that malingering is a trait that manifests itself across most activities and situations. More properly viewed, malingering is situation specific and depends upon various factors. Often malingering is an appealing solution to individuals who experience social, personal, or economic problems (Lezak, 1983). It is a strategy used in different areas, but predominately being within the legal system. Many times people who feign various symptoms tend to find themselves in a difficult situation and feel it is the only way to resolve their problems.

As indicated above, the type of behavior that a person feigns depends on different sets of circumstances. This is, all cases of malingering will not display the same types of behaviors. Thus, a person who is feigning a back injury will show that he/she has trouble lifting objects, bending over etc. but will demonstrate normal attention and memory. While this example may seem simplistic, almost trite, it is not so obvious which cognitive

dimensions will be feigned by a person with a head injury. These examples amply illustrate that it is impossible to develop a malingering test that will be appropriate across all situations. For example, within the area of forensic psychology a variety of different tests may be used to determine if an individual knows the difference between right and wrong. However, the procedures used for that situation are far different from those that are used to determine if a person is faking those cognitive symptoms that occur with a concussion.

It is also important to distinguish between malingering and psychiatric diagnoses, such as factitious disorders (Munchausen and Ganser syndromes) and anxiety-based disorders (somatoform and conversion disorders). With factitious disorders, the individual intentionally produces symptoms but does so to assume a sick role and is not motivated by external incentives. Conversely, an individual with a somatoform disorder exhibits similar symptoms, however these symptoms are not under voluntary control. In other words, two components, falsification and intentionally, distinguish malingering from that of the other disorders (Faust & Ackely, 1998). In the DSM manuals (APA, 1994), the term “Malingering” lays on a different diagnostic axis than conversions and somatoform disorders. This diagnostic difference makes it clear that malingering is not a psychiatric diagnosis. Below are further detailed the clinical features and degrees of malingering. For further information on the terminology of malingering the reader is directed to the following readings: Miller (1996), Nies and Sweet (1994), and Rogers (1988). Psychologists want to feel confident in distinguishing between individuals with a conversion disorder and those who believe they have impairment, from those who are consciously faking a deficit.

Lipman (1962; as cited in Nies & Sweet, 1994) was the first author to explain malingering as a behavior with the following characteristics: invention, preservation, exaggeration, and transference. Invention refers to the fabrication of symptoms or deficits; preservation refers to symptoms that were at one time present but no longer exist or have subsided; exaggeration refers to symptoms that are worse than they actually are; and transference refers to symptoms that are present but incorrectly and intentionally attributed to a particular injury.

In addition, Rogers (1988) stated that there are different degrees of malingering: mild, moderate, or severe. Mild malingering is defined as an exaggeration of symptoms and consequently only plays a small role in differential diagnosis. A moderate level of malingering is represented by an individual who either fabricates or exaggerates symptoms in an attempt to present oneself as more disturbed than one actually is. Finally, a severe malingerer would be defined as someone who blatantly fabricates or exaggerates symptoms to the point of absurdity.

Although malingering can occur within various disorders, this paper examines malingering from a neuropsychological prospective, primarily focusing on the aspect of mild head injuries. A mild traumatic brain injury can be defined by some of the following criteria (MTBI; Alexander, 1995; Trueblood & Schmidt, 1993; Zielinski, 1994):

1. External trauma to the head is necessary.
2. Loss of consciousness less than 20min, however, unconsciousness does not necessarily need to occur.
3. Glasgow Coma Scale of 13 to 15

4. Post traumatic amnesia less than 24 hours.
5. Neuroimaging techniques is negative.
6. Deficits in attention, information processing, and interference.

As a side note, individuals suffering from a moderate TBI would fall into the Glasgow Coma Scale of 9-12. For further information on the neuropathology, measurements and clinical management of traumatic brain injuries the reader is referred to such articles as Alexander (1995) and King (1997).

The most common symptoms reported by MTBI patients are attentional difficulties and memory complaints (Binder, 1986; Zielinski, 1994). Accordingly, these symptoms are the most commonly feigned. Individuals who uncharacteristically maintain cognitive difficulties from a mild TBI beyond the 90-day expected recovery period are considered to have atypical symptoms. Slick, Sherman, and Iverson (1999) stated that it is these atypical symptoms that raise the question of malingering. In many cases conventional radiological measures of brain damage are negative, however, this does not necessarily imply that the problem is of a psychological nature.

In the past, neuropsychologists have relied on clinical test batteries, experience, and knowledge about the brain behavior relationship to detect malingering. Further improvements have been made along the way but it seems no one test can measure malingering with 100% certainty. There has been an increasing interest in the reliability and validity of being able to correctly detect an individual attempting to malingering. Thus, to ensure the legitimacy of a claimed deficit (Vallabhajoula & van Gorp, 2001), the role of the neuropsychologist in being able to accurately detect intentional under performance is of the utmost importance. In the absence of radiological evidence, we rely heavily on a

neuropsychologist's testimony. The neuropsychologist is an important expert in court cases for TBI in that they provide non-radiological evidence in the form of neuropsychology test battery scores. These tests, in turn, must be valid to rule out other diagnostic possibilities such as malingering.

Why is the Study of Malingering Important?

Studying malingering is imperative for several reasons, all of which relate to the issue of validity. The first reason is that we need to distinguish between those individuals who are genuinely cognitively impaired from those who are not. The problem with making this differentiation is that current neurodiagnostic methods do not always detect the subtle deficits that may have occurred (Binder, 1986; Guilmette, Sparadeo, Whelihan, & Buongiorno, 1994). Thus, it is important that we devise specific psychological and neuropsychological tests that are capable of distinguishing between feigned and genuine impairments.

It is of the utmost importance that clinicians maintain impartiality in servicing both the patient and the judicial system (Miller, 1996). One must be mindful that tests which detect intentionally reduced effort are used not only for the purpose of catching those who are malingering but also for ruling out those who are not, in order that appropriate treatment and compensation can be amended (Bianchini, Mathias, & Greve, 2001; Slick et al., 1999). Neuropsychologists are commonly asked to differentiate between those who are faking and those who are not. When approaching a diagnosis, one must be wary of answering such a definitive question.

It is important to study malingering because if it goes undetected society pays in terms of invalid claims through higher insurance premiums or tax costs. On the other

hand, a genuinely injured individual may be denied certain medical benefits to aid in recovery and support, as well as suffer emotionally due to an incorrect diagnosis (Bianchini et al., 2001; Strauss et al., 2000). Services may be withheld from those in need as a result of inaccurately classifying someone as possibly malingering (Franzen, Iverson, & McCracken, 1990). However, failure to detect malingering needlessly provides services and indirectly keeps treatments from someone else.

Secondly, we examine malingering in order to maintain credibility within the psychology profession. Interpretation of test results is based on the assumption that the patients are putting forth their best efforts. The patient's cooperation is key to establishing the test's validity (Nies & Sweet, 1994). The integrity of all test results is compromised when the individual does not put forth his/her best effort (Binder & Willis, 1991; Hiscock & Hiscock, 1989; Rogers, 1988).

A final reason why we look at intentional under performance is to study the topic of malingering itself. Why do people malingering? What is the motivation? Are certain profiles more likely to malingering than others? The diagnosis of malingering should be treated as any other 'disorder'. Furthermore, we should look at the pragmatics behind its occurrence and investigate ways of preventing and treating malingering (Slick et al., 1999).

History

Up until the mid 1980's, most psychologists believed that the detection of malingering could simply rely on qualitatively examining how a patient performed on neuropsychological tests. Examples of potential indicators of malingering are as follows:

1. A high number of 'I don't know' responses.

2. Getting hard items and missing easy ones (Slick et al., 1999; Ruff, Wylie, & Tennant, 1993).
3. Approximate answers or near misses.
4. Test results that don't make 'neurological sense'. For example, the patient does well on Trails B but poorly on Trails A.
5. Inconsistencies between and within neurological tests that assess the same cognitive area (Iverson & Binder, 2000; Larrabee, 1990; Ruff et al., 1993; Slick et al., 1999).
6. Poorer performance on recognition tests than on recall tests (Bernard, Houston & Natoli, 1993; Faust & Ackley, 1998; Nies & Sweet, 1994).
7. Examination of the serial position effect (Slick et al., 1999; Strauss et al., 2000).
8. Discrepancies between reported and observed everyday activities (Iverson & Binder, 2000). For example, if a patient complains of word finding problems but yet is able to articulate him/herself well while testifying (Larrabee, 1990); complains of memory difficulties yet the patient can find his/her way back to the clinician's office without aid.

Trueblood and Schmidt (1993) examined the characteristics of malingering and under performance on a selection of neuropsychological measures. Their study used 81 malingering patients selected from a sample of 106 MTBI outpatients who had external incentives (such as seeking benefits or were involved in litigation). Malingering patients were identified by obtaining a significantly below chance score on a symptom validity test (test procedure is discussed later). In addition, a Control group matched to age,

education, and gender, consisted of patients involved in motor vehicle accidents that did not meet all MTBI criteria and were not seeking external incentives were used. All participants were measured on intentional poor performance using 14 indicators, some of which are listed above. They found that many of the suggested invalid indicators were either not present or, if so, occurred rarely. This result supported the notion that many of these signs were inadequate for detecting malingering. Furthermore, if such indicators are relied upon for detecting malingering it could result in overconfidence in the validity of the neuropsychological data.

Perhaps the most dramatic example as to why a reliance on qualitative symptoms will fail to detect malingering is provided by Faust, Hart, Guilmette and Arkes (1988). Their study showed an extremely low detection rate of malingering among neuropsychologists when they were provided with only test scores. In two separate studies, Faust et al. (1988) examined the ability of neuropsychologists to detect malingering in adolescences and children. In their first study they instructed three adolescents, aged 15 to 17 years, to intentionally under perform on the Wechsler Intelligence Scale and the Halstead-Reitan neuropsychological battery (HRNB). They were initially paid \$15 and told that if they went undetected they would be paid an additional \$5. Neuropsychologists were given a history that the adolescent patients sustained a brief loss of consciousness, an unremarkable CT head scan, a report that school grades had deteriorated, and that the patients were forgetful. This information, together with test scores, was mailed to 300 neuropsychologists. Their study found that 80% of 113 neuropsychologists who returned protocols diagnosed cortical dysfunction as the reason for the abnormal test scores. In a second part to the first study, pairs of

protocols (2 malingerers and 2 head injured patients randomly selected) were given to 26 neuropsychologists. They were informed of a 50% base rate of malingering and were asked to diagnose the cases. However, even when judges were armed with this base rate information, inaccurate diagnoses remained; there was a 10% false positive rate and only a 43% true positive rate.

A second study by Faust et al. (1988) used children aged 9 to 12 and examined the same prospects of detecting malingering as in their first study. Results of their study indicated that 93% of the neuropsychologists misattributed the cases to either cortical dysfunction or functional factors. In addition, the study went a step further by asking the judges how confident they were in their diagnosis and found three quarters of them indicated they were moderately or highly confident in their decision. In combining these two studies, Faust et al. showed any individual, regardless of age, has the potential to simulate malingering. They concluded that qualitative signs were not useful and do not occur frequently enough for an accurate detection rate. This conclusion was further supported by Trueblood and Schmidt (1993, discussed on the previous page).

In challenging Faust et al. findings, Bigler (1990) suggested that the low detection rate in their study was directly related to neuropsychologist's low suspicion and the tendency of adolescents to malingering. Additional criticisms included that there were some neuropsychologists who were not experienced with the HRNB, some information was withheld, behavioral observations could have aided the practitioner in their decision, no background information was given, and judges were not told to look at malingering as a possible explanation of test abnormality (Bigler, 1990). In other words, it was argued that relying solely on test scores did not allow for a complete and representative opinion.

In spite of these criticisms, the fact remains that not one neuropsychologist in either study detected malingering. Thus, although some psychologists thought that the Faust et al. study was not representative of a real decision-making process, their results provided a startling picture that a clinician's confidence in diagnosing malingering solely on the basis of test scores was gravely overestimated. Needless to say, Faust et al.'s hallmark study was a major stimulus in sparking further interest in trying to find procedures that would improve the detection of malingering cognitive deficits.

Further exploration into the issue of accurately diagnosing malingering using neuropsychological test data was examined by Trueblood and Binder (1997). Their study used clinical adult malingerers and TBI participants. Four individuals were judged to be malingering cognitive deficits on the basis of the following criteria: below chance test scores on the Portland Digit Recognition Test or the Digit Memory Test, and other supporting data including some supplied by surveillance equipment. In addition, all four individuals were involved in litigation and/or seeking workers compensation. Two TBI individuals, not involved in litigation, were also used in the study. All participants completed the extensive neuropsychological test evaluations. Each judge received summaries, which included a case history and litigation circumstances, on two participants (either two malingers or two head injured individuals). Scores from malingering tests were included in the summaries for only two of the four malingering cases, resulting in a higher confidence of a correct diagnosis by the judges. The error rates for detecting malingering ranged from 0% to 25% across all four cases of malingering. These results were superior to those found in the Faust et al. (1988) study. It may have been that using actual clinical cases, as opposed to simulators, provided a

more realistic situation. Additionally, payoffs for those involved in legal cases are more substantial, personality factors such as being a risk taker may also be indicative of those attempting to malingering, and finally individuals malingering may have a greater knowledge base of deficits that occur with a head injury (Trueblood & Binder, 1997).

Furthermore, results of this study may have been affected by the high degree of notoriety of the Faust et al. (1988) study, thus increasing clinician's knowledge and sensitivity to the issue (Trueblood & Binder, 1997). However, Trueblood and Binder acknowledge their findings cannot be generalized due to several factors: First, their sample size was too small. Second, their study only used obvious malingerers perhaps making them easier to diagnose. Third, less than a quarter of the forms requesting participation in the study were returned, thus only those clinicians that were confident in their ability to correctly diagnose may have been involved in the study. Fourth, neuropsychologists were told in the covering letter that malingering was a possible diagnosis and thus this information may have tipped them off. Finally, judges were given limited information that was not representative of a true clinician's situation.

More recent exploration of how traditional neuropsychological batteries may aid in detecting malingering have been examined. In support of earlier findings, Van Gorp et al. (1999) also found that examining patterns of neuropsychological test performance was an unreliable method for detecting malingering. Collectively, all of the studies described above support the notion that malingering cannot be reliably detected through examination of neuropsychological test scores. Thus, based on these past studies it was imperative to establish quantitative measures for detecting malingering.

Early Malingering Tests

As previously mentioned, neuropsychologists were at one time confident in their ability to detect malingering using neuropsychological indicators and thus there seemed no need to use other detection means. However, in light of the research mentioned above there was a need to employ a test that was directly aimed at detecting intentional poor performance. Some of these tests are described below.

Dot Counting Test (Rey, 1941): One of the earliest tests specifically aimed to detect malingering was the dot counting test developed by Rey (1941, cited in Lezak, 1983). The test consists of a set of cards with dots (varying anywhere from 4 to 30 dots) that are either grouped or ungrouped. Grouped dots are arranged in specific patterns such as a diamond or square. These cards are randomly presented to the examinees; their task being to count the number of dots displayed. The reasoning behind this procedure is that one's ability to count dots is simplified when dots are grouped as opposed to ungrouped. Thus, test results should show a faster response for grouped dots than ungrouped. The test is based on the theory that those individuals who are attempting to malingering will have a slower response time for grouped than ungrouped cards. Unfortunately, the test's inability to distinguish between controls, malingerers, and those with brain injury has been shown to be an unreliable way to detect malingering (Bianchini et al., 2000; Bink, Gouvier, & Waters, 1997; Greiffenstein, Baker, & Gola, 1994).

Rey Fifteen Item Test (Rey, 1964): In an attempt to improve measurements of detection Rey (1964) designed the 15-item memory test, which is still somewhat popular today. This test consists of one card containing 15 item grouped letters, numbers, and symbols: A,B,C, a,b,c,1,2,3,O,□,Δ,I,II,III. The examinee is shown the card for 10 sec and

asked to remember all the items. Following this presentation the participant is then asked to write down what he or she remembers. Emphasis is placed on the number 15 in an attempt to make the task appear more difficult when in reality there are only three concepts to remember (Nies & Sweet, 1994). Since there are really only three concepts to remember (letters, numbers, symbols), the assumption is that one attempting to malingering will obtain a low score.

Varying results have been found with the 15-item test. A study by Davidson, Stuffield, Orenczuk, Nantau, and Mandel (1991) compared various subpopulations (i.e., chronic pain patients, spinal cord injured patients, brain injured patients, simulators, and controls). They found a significant difference between the controls and the simulators. However, when comparing the simulators to brain injured or chronic pain patients no differences were detected. Another study by Guilmette, Hart, Giuliano, and Leininger (1994) showed patients with moderate-to-severe brain injuries performed worse than depressed participants, controls, and simulators. They found that those attempting to simulate memory deficits actually obtained scores that would not raise suspicions of malingering.

Greiffenstein et al. (1994) compared the performance of the Rey 15-item test with individuals having traumatic brain injuries to probable malingerers. Results of their study found that probable malingerers scored well below those with brain injuries. However, participants in the brain injury group who were involved in litigation were also known to have a potential impact on lowering test scores. Another weakness to their study was lack of a Control group.

A study conducted by Schretlen, Brandt, Krafft, and van Gorp (1991) examined the 15-item test's ability to detect malingering by administering the test to individuals asked to feign a mental disorder and to individuals who actually suffered from a particular deficit or disorder. They found 27% of the participants that actually suffered from either a mental or neuropsychiatric disorder were misclassified as malingerers and less than 15% of individuals attempting to simulate malingering were identified. These results do not support the Rey 15 item test as being a valid method of detecting malingering.

Studies have examined the Rey 15-item test by using varying cut-off scores. Guilmette et al. (1994) examined the differences in sensitivity with cut-off scores of 7, 8 and 9. Their study used moderate to severe brain damaged individuals, depressed patients and controls who were asked to simulate memory impairment. Using a cut-off measure of 7 revealed a sensitivity rating of 5%, a cut-off score of 8 showed a sensitivity of 15%, and a cut-off score of 9 reported a sensitivity of 25%. Nonetheless, regardless of the cut-off score used, low sensitivity still remained.

Although easy to administer and score, the 15-item test is too simple and may tip off its primary purpose to the examinee (Davidson et al., 1991; Schretlen et al., 1991; Vickery, Berry, Inman, Harris, & Orey, 2001). In addition, some studies have been unable to differentiate between controls and simulators (Bernard, 1990; Schretlen et al., 1991) while others have reported sensitivity to genuine impairments (Davidson et al., 1991). Because of the test's lack of reliability and transparency, it is judged to be an inadequate measure of detecting malingering.

Symptom Validity Tests (SVT)

Prior to the 1990's the dot counting and the Rey 15 item proved somewhat useful in detecting malingering, however an improved method was still needed. This set the stage for the symptom validity test (SVT: Pankratz 1979, 1983) which added prediction power and provided a move towards a less qualitative way of detection than its predecessors.

Symptom validity testing relies on the laws of probability and allows one to empirically observe if an individual is malingering. For example, one of the first studies to use SVT in detecting malingering was by Binder & Pankratz (1987). The study involved a woman who insisted she had memory deficits. Neurological evaluations were normal, thus to investigate whether her complaint was legitimate, they first showed her either a black pen or a yellow pencil. The object was removed from her sight once she acknowledged recognizing the stimulus. The patient was asked to count aloud from 1 to 20 (distraction task) and then asked to recall which object she had just previously seen. One hundred trials were administered to the patient. If the patient were performing to the best of her ability and had a genuine perceptual deficit, then she would be able to correctly identify the correct stimuli 50% of the time, by chance of guessing, because she was unable to remember the stimuli. However, the patient only obtained 37% correct score, a below chance performance, an indication of intentional under performance.

Today's SVT procedures consist of a two-item, forced-choice test to detect malingering. An initial stimulus target is presented and shortly thereafter the same stimulus is shown together with a foil. The examinee is presented with a two forced-choice discrimination task, the stimulus target and the foil. A wide selection of target

stimuli can be presented such as digits, words, or even pictures. For example, if the participant is presented with the word “tree” as the study stimulus and shortly after is simultaneously shown the words “tree” and “bush” as the choice stimulus, they would be forced to choose the word they had just previously seen. SVT procedures attempt to improve the sensitivity of detecting malingering while impacting little on its specificity. In addition, the SVT procedure is very flexible and adaptable because any stimulus may be used to suit the clinician’s assessment needs (Bianchini et al., 2001).

Two scoring methods are used in diagnosing malingering, a below chance score and an empirically determined cut-off score. A below-chance performance is indicative of a patient responding actively with a bias due to either malingering response style or not cooperating with test procedures (Bianchini et al., 2001; Fredrick & Foster, 1991; Larreebe, 1992). As scores decrease there is an increased likelihood that the individual is intentionally selecting wrong answers (Millis & Volinsky, 2001). Using below chance scores often result in making false negative conclusions (Martin, Smirolodo, Strafford, & Gouvier, 1994) due to malingering sophistication. In other words, using a below-chance score to detect malingering is uncommon as it often results in low sensitivity. Therefore, we tend to use cut-off scores that are based on an ‘empirical floor’ that is normed-referenced, a clinical population where the lowest score obtained in the group represents the cut-off score (Bianchini et al., 2001). If a test is scored out of 20 and the lowest score obtained in the clinical reference group (i.e., individuals with TBI) is 17, this score would be used as a cut off measure. Any score below it would suggest intentional under performance. Thus, if one is attempting to malingering a brain injury and obtains a lower

score than that of the lowest score obtained in the reference group, it is likely that the individual is malingering.

The SVT technique relies on the assumption that intentional feigners will overestimate the level of impairment that a TBI patient will have on a recognition memory task (Fredrick & Foster, 1991; Larreebe, 1992). That is, it is based on a floor effect, where an individual malingering will fail a task in which even grossly impaired individuals will do well (Trueblood & Binder, 1997). SVT tests are designed so that, in fact, little memory is required and patients with genuine impairments do well (Bianchini et al., 2001; Guilmette et al., 1994).

Building on the SVT procedure, Hiscock and Hiscock (1989) came up with what is currently known as the digit memory test (DMT). This test consists of 72 trials split into 3 blocks (24 items per block). On each trial a string of five-digit stimulus is presented followed by a two-choice recognition test. For example, the examinee is shown '23649' and after a brief delay the person is shown 2 numbers e.g. '23649' and '78125'. The examinee is then asked to choose the number previously seen prior to the delay. The incorrect number never has the same first digit as the correct number thus the test is quite simple. The delay between target presentation and recognition increases between blocks with the first block having a delay interval of 5s, the second block 10s, and the third 15s. The reason behind presenting a string of five-digit numbers and increasing the delay intervals between blocks was two fold. Firstly, presentation of five-digit numbers increased the plausibility of a memory test, thereby decreasing the transparency of a test aimed at detecting intentional underperformance. Secondly, increased intervals were incorporated to give the impression the task became harder when

in fact it did not, thus the test appeared more difficult than it was (Bianchini et al., 2001; Prigatano & Amin, 1993). Many of today's SVTs or tests that follow a SVT format are based on the framework of the DMT.

Binder (1993) looked to further improve the DMT, and developed the Portland Digit Recognition Test (PDRT). The PDRT is similar to that of the DMT (5 digit numbers presented visually followed by forced choice procedure) except that within the delay interval an interpolated activity occurs whereby the person is required to count backwards from a predetermined number. Thus, more so than the DMT, the test appears to be more difficult than it is which tends to make it sensitive to individuals with TBIs (Bianchini et al., 2001; Vickery et al., 2001). Consequently, since TBI patients obtain a low score on this procedure, it is difficult to distinguish between TBI patients and individuals who are malingering TBI symptoms.

Although the symptom validity techniques showed considerable promise and improved upon past tests aimed at detecting malingering, methodological variations of the test offer more adequate alternatives.

Methodological Variations of the SVT

Iverson et al. (1991) designed a variation of the SVT that consisted of a 21-word list containing nouns. Each word is presented orally to the examinee. After the entire list is read the examinee is asked to recall as many words as possible and then is given a two-choice recognition test. Their study used a malingering group, a control group and a TBI group; correct classification rating of 65%, 100% and 100% were found respectively. A potential drawback of this test is there are few trials. With this drawback kept in mind

and the addition of more trials, the 21-item test holds promise as an alternative test in detecting malingering (Vickery et al., 2001).

Another variation of the SVT technique by Green, Iverson, & Allen (1999) assesses both verbal memory and response bias by using what is called the Word Memory Test. The examinee is presented with 40 common words in a paired association format (close semantic association such as woman/man and subtle semantic linkage, such as tree/lake). After the list is administered twice, a variety of tasks varying in difficulty are given: immediate force choice recognition, delayed recognition, multiple choice recognition task, paired association recall, cued recall and free recall. Each task is then evaluated in terms of the examinee's effort. A literature review on the WMT revealed little empirical evidence on the test.

The Test of Memory Malingering (TOMM, Tombaugh, 1996) is yet another methodological variation of the symptom validity tests. The procedure of the TOMM is as follows: the examinee observes 50 pictures (study trials) for 3 seconds each. Immediately after, a two forced-choice recognition procedure is employed whereby 1 old and 1 new picture is shown. The participant chooses one of the pictures shown and feedback on correctness of the response is provided. This procedure is used twice. Fifteen minutes after the second administration, a final 2-choice recognition test is given.

There have been several validation studies to the TOMM: cognitively intact and neurologically impaired patients, university undergraduate students who served as either controls or simulators, TBI patients who served as either as controls or simulators; litigation vs. nonlitigation participants, and depressed vs. non depressed patients. These studies have shown a 90% correct classification rating for cognitively impaired, 98% for

TBI, 100% for depressed individuals, and 77% for TBI patients involved in litigation. Using undergraduate students as simulators or controls, Rees et al., 1998 found a 95% sensitivity rating and a 100% specificity rating. The TOMM has extensive norms from various subpopulations (Tombaugh, 1997), and is one of the most popular methods used to detect malingering.

Computerized techniques aimed at malingering detection

Computerized tests not only offer the added benefit of response latencies but also allow for automatic scoring procedures that are more efficient and less prone to error (Bianchini et al., 2001). However, the use of computers does have its downside, for example, individual performance may differ due to varying familiarity with the computer (Bianchini et al., 2001). Moreover, use of a computer requires a degree of cognitive, visual spatial and motor skills and may not be appropriate for certain populations (Rose, Hall, & Szalda-Petree, 1995). Nevertheless, intuitively, results from a computerized test appear to have greater face validity than non-computerized tests.

Over the past few years, the use of a computer seems to be a popular way of studying and detecting poor motivation. For instance, the Victoria Symptom Validity Test (VSVT; Slick et al., 1994) is a computer-administered, forced-choice, recognition procedure of the DMT with the addition of easy and difficult trials in the hopes of increasing sensitivity and detecting motivational troubles. Easy items are paired with foils that share no common digits ('34028' vs. '56197'); hard items invert the middle two digits while the remaining digits stay in the same sequence ('46923' vs. '46293'). Thus, knowledge of the correct answer for hard items requires the examinee to correctly recall the order of the middle digits (Doss, Chelune, & Naugle, 1999). Studies have shown that

both simulators and compensation seeking participants have a greater number of errors and longer response times than control participants. This statement holds true more so for hard items rather than easy items (Bianchini et al., 2001; Strauss et al., 2000). Doss et al. (1999) found stable performance on the VSVT from various clinical patients who differed in cognitive ability. More importantly, results also signified that those involved in litigation performed more poorly than those who were not involved in seeking compensation.

The Test of Memory Malinger – Computerized gives the added benefits of gathering response latency information. Rees, Tombaugh, Gansler and Moczynski (1998) found longer response latencies for a simulator group than a control group. They hypothesized that this result was due to increased processing time. Not only does the ‘malinger’ participant need to determine which answer is correct but he/she was also faced with a decision of whether to answer with truth or deception, thus resulting in a longer reaction time. However, this study lacked a TBI group for comparisons purposes.

Some critical issues in the study of malingering

Motivated by its extreme importance within the legal realm, there has been increased interest over the past two decades in developing better ways of detecting malingering. Originally qualitative signs were employed, but without much success. Then the emphasis shifted toward development of specific malingering tests. Currently there has been the realization that no single approach will provide the type of evidence needed to detect malingering and clinicians have embraced a more omnibus approach. Within this context, Nies and Sweet (1994) proposed the following strategies for improving the detection of intentional under performance: use of specific malingering

tests, assessing patterns of performance, evaluating inconsistencies across and within tests scores (eliminating fatigue, changes in motivation etc), gathering of a patient's clinical history, and seeking independent information that either confirms or refutes a patient's behavior.

However, the major burden of evidence still resides in quantitative measures such as those provided by specific malingering tests. In reviewing tests designed for detecting malingering, several issues still remain. The issue of generalization, coaching, universal guidelines, base rate information, and the issue of labeling also remain problematic.

One of the biggest questions that haunt researchers is the issue of external validity: can we generalize a simulator's performance to an individual who is "malingering" in an actual court case or on a clinical neuropsychological assessment? In other words, can we generalize results from an artificial setting to the real world? Bianchini (2001) suggests comparing a likely-to-malinger compensating group to a not-likely to malingering compensation group to improve the validity of a test aimed to measure feign deficits. However, it is often hard to determine which, if any groups fall into the category of 'likely-to-malinger'. The best fit for this population -those involved in litigation. (See Faust & Ackley, 1998; Schacter, 1986 for a detailed discussion on generalization).

Second, are there differences between participants who are coached on how to malingering and individuals who are not coached? Opinions in this area vary (see Ben-Porath, 1994; Bianchini et al., 2001; for discussions on this topic). Just because we ask a participant to malingering doesn't mean they can. An additional question to consider is

“Should we evaluate strategies used by simulators?” The answer to this question is “most definitely”, as the added information can only enhance understanding of malingering.

Third, there is no universal set of instructions for asking participants to simulate symptoms indicative of malingering and thus instructions differ from one study to the next (Bianchini et al., 2001). Also, simulators may not follow the instructions to malingering properly. On the other hand, simulation studies for detecting malingering prove useful in showing patterns of neurological deficits similar to patients with actual TBI (Andrikopoulos, 1994) and aid with the development of new detection methods (Strauss et al., 2000).

Fourth, studies of malingering have an inherent limitation – there is no base rate and thus no error rate can be established (Faust & Ackley, 1998; Slick et al., 1999). Base rate information is not available in a real clinical situation, as the prevalence of malingering in relation to head injury remains uncertain (Trueblood & Binder, 1997). One can most certainly agree that it is rare that an individual will readily admit to malingering. Moreover, genuine impairment may coincide with feigned deficits, thus distinguishing between the two becomes imperative but unfortunately problematic (Faust et al., 1988). In the absence of base rate information what then can be done to improve validity of a test? One improvement is within subject testing or multiple testing times. Another improvement is the collection of data from multiple sources. Finally, gathering information regarding the test itself and the disorder you are attempting to detect is also important.

Fifth, does the use of incentives change the outcome of performance? Most studies do not have a comparable payoff to that of being compensated through an

insurance agency or claims court. The motivation for a volunteer, which are usually psychology undergraduate students, is either course credit or a nominal fee. This small incentive vs. a large reward for a personal injury case is difficult to compare, as one clearly would put more time and resources into going undetected when a large sum of money is involved. (see Andrikopoulous, 1994 for more details on using incentives).

Sixth, studies aimed at detecting feigned memory deficits rely on the layman's view that memory is a unitary process rather than a set of interrelated processes (Aubrey, Dobbs & Rule, 1989). Some feel that although patients have had some previous knowledge of TBI clinical symptoms (from symptoms they used to have) it appears they cannot apply that knowledge in a believable manner (Nies & Sweet, 1994; Wiggins & Brandt, 1988). They are unable to self-monitor and implement their knowledge appropriately. Assertions have been made that individuals with personal experience have no better knowledge about head injury than those without experience (Gouvier, Prestholdt, & Warner, 1988; Martin et al., 1994). A study by Rees et al. (1998) examined the ecological validity of using university simulators to depict malingering by comparing their performance to that obtained by TBI participants who were asked to "malinger". The study showed that TBI individuals performed similar to university simulators, and they used similar strategies in their attempt to malinger. Therefore, based on these results, we can confidently assume that university simulators may present realistic TBI symptoms when attempting to malinger. It is thus of utmost importance that a patient's knowledge of not only the disorder but also the test be investigated (Zielinski et al., 1994).

Finally, with an increasing ability to detect malingering, comes the problem of ‘labeling’ (Pankratz & Erickson, 1990). To brand someone as being a “malingerer” has several drawbacks. First, we can never be one hundred percent sure the person is faking. Second, the person is likely to be deprived of not only money but also medical treatment. Third, the label may remain with them for life and thus possible future impairments that are, in fact genuine, may be questioned with suspicion and ultimately denied treatment. As psychologists, we tend not to label someone as a malingerer and instead allude to the fact that the patient’s performance is inconsistent with a certain profile, or seems to fit a pattern of malingering, or didn’t put forth their best effort.

Why Further Test Development is Necessary

As indicated above, one of the current trends in the detection of malingering is to provide converging evidence through the use of different tests. At the present time, most of the tests designed to detect simulation of cognitive symptoms have used the SVT methodology. This over reliance on a single paradigm tends to reduce the breadth of psychometric evidence that is typically used for the detection of malingering. This reliance on existing tests is even more problematic in view of the fact that information about these tests is readily available on the Internet. A recent article in the Boston Globe (Barry, 2002) stated the real possibility of malingering individuals being able to go undetected due to their searching for information and advice via the Internet on how to respond to specific malingering tests.

Thus, new tests or revisions of older test must be researched and implemented as knowledge about currently used tests become known (Nies & Sweet, 1994). In addition,

the SVT paradigm is becoming well used and in turn well known; there is an increasing need to find a new paradigm so that individuals attempt to 'beat the system' are detected.

So, what can be done to improve upon a test's ability to accurately detect malingering? Several factors lead to a superior measure of malingering (Tombaugh, 2001). First, the test should be sensitive to malingering, meaning it should be able to differentiate an individual who is malingering from one who is not. Second, the test should be insensitive to all other variables such as age, education and cognitive deficits. Third, the test should have high face validity, it must appear to be part of a legitimate neuropsychological battery otherwise the malingerer may be suspicious of what the test is attempting to measure. Finally, the perceived difficulty should be greater than the tests actual difficulty. Therefore one must keep these criteria in mind when devising a valid malingering test.

With these improvements in mind, better methods of detection in turn can aid the judicial system in determining scientific admissibility of evidence. A set of specific parameters, known as the Daubert guidelines, was devised to increase reliability and validity of an explicit tests methodology (Bianchini et al., 2001; Vallabhajosula & van Gorp, 2001). Contrived in the form of questions (Tombaugh, 2001), these guidelines are as follows: Question 1. 'Has the technique been tested or is subject to being tested?' 2. 'Has the test been subjected to peer review and publications?' 3. What is the known or potential rate of error in applying this procedure?' and finally 4. 'To what extent has the test been accepted in the relevant scientific community?' If supportive answers are produced for these guidelines the test will generally fare well within the 'malingering'

test realm. This in turn leads to an increased dependency on tests that are valid and reliable.

In summary, test improvements are imperative due to the increased access and ability to uncover how tests aimed at detecting malingering work. Additionally, there are intensified demands on our legal system to accurately determine individuals who are malingering. The use of the computer has increased substantially over the last decade due to its ability to accurately retain information, gather data in a more efficient way, and provide us with additional information such as latency times. A test must be reliable and valid to accurately detect intentional underperformance. Therefore, exploring how reaction times, via the computer, may be used as a new indicator of detecting malingering seems warranted.

Reaction Time Measures of Malingering

Reaction time (RT) measures are well documented to be sensitive to deficits in information processing capacity that occur in TBI (Collins & Long, 1996; Van Zomeren & Brouwer, 1994). In addition, Resnick (1988) speculated that individuals attempting to malingering a head injury should have longer RTs than those obtained by TBI patients. This speculation was based on the premises that (1) simulators will overestimate the amount of time that TBI patients require to make a response and (2) the decision processes involved in formulating and executing a “malingered response” will unduly increase response times. However, as the following review will reveal, the evidence showing the utility of RTs to detect malingering is inconsistent.

Latency Measures in Conventional Malingering Tests

Besides looking at the number of correct responses in conventional malingering tests, many studies have also found that latency information may aid the examiner in detecting intentional underperformance. Rose et al. (1995) was the first to look at the usefulness of employing RTs with conventional malingering tests as a way to detect poor motivation. They compared the performance of moderate head injured patients to those of controls and malingerers on the Portland Digit Recognition Test – Computerized. The Control and Malingering groups consisted of university students who received course credit for participation. Each group consisted of 30 participants and a cut-off of 39 for the PDRT-C was used. The malingering participants were given instructions and told they would receive extra course credit if they successfully deceived the examiner. The reading of the instruction and test administration occurred on the same day. No other tests were given.

The PDRT-C required the participant to press either 1 or 3 on the numeric keypad for responses to the left or right items. Volunteers were asked to keep their fingers on the keys at all times. During the delay period, instead of counting backwards, the participant was asked to press the space bar whenever a specific target appeared on the screen. However, one may question the complexity of using three different key responses: 1, 3, and the spacebar. Administration was as follows: a fixation point was presented for 2s, then a target was displayed for 5s followed by the vigilance task during the delay period. At the end of the delay interval, the two stimuli were presented and finally feedback.

Expectations were that the malingering group would obtain a longer RT than both the TBI and control groups. They found that clinical head injury patients had longer RT

than both the control group and the malingering group, with the control group having the shorter RTs. Thus, the assumption that malingering can be detected via longer response latencies was not supported. This finding can be explained, at least in part, by the use of moderate-to-severe brain injured individuals. Due to the severity of their impairment, these individuals are known to have longer latencies than those with mild head injuries who tend to have shorter RTs.

When Rose et al. looked at a discriminant function analysis that took into account both RT and number of items answered correctly, they found sensitivity improved from 64% to 74% and specificity from 89% to 100%. These results indicated that although response latencies per se may not be used to detect malingering, they can improve the accuracy of a tests in detecting malingering.

Inspection of latency scores on the Test of Memory Malingering-Computerized (TOMM-C: Rees et al., 1998) found differences between a simulator and control group. Forty participants, consisting of both university students and volunteers from the community, were randomly assigned to either a control or simulator Group. They found that RTs for correct responses decreased over trials for both groups, however, the simulator group had a longer RT for correct responses. Unfortunately, this study did not provide any data from a TBI group for comparison purposes. In spite of this, RT scores on this test were valuable in helping the examiner detect malingering.

A study by Strauss et al. (2000) looked at RT differences on the Victoria Symptom Validity Test (VSVT) between a Control and a Simulator group (undergraduate students). The simulator group was given a set of instructions telling them to demonstrate a realistic head injury impairment for litigation purposes. In addition, they

were told the amount of money they obtained from the lawsuit would depend on their performance and obvious faking would be detected. The VSVT was embedded within other neuropsychological tests. All test were given three times, at weekly intervals.

Results showed that response latencies on the VSVT for the Malingering group were slower than those for the Control group. These latency differences were also more pronounced with hard items rather than easy items. When looking at the RT for hard items, a prediction rate of 72.2% for controls and 70.8% for malingerers was found. The variability on RT was more pronounced for the malingering group, particularly for the hard items. Results of repeated assessments indicated a slow improvement for both the control and malingering groups with the latter group still obtaining slower response times than the control group. By the third session, only RTs on hard items were significant. The authors concluded that one testing session was sufficient in detecting malingering but that multiple testing sessions can aid in detecting discrepancies of intraindividual variability. Thus, their findings support using latency scores in detecting intentional underperformance. However, their study is flawed, as was the Rees et al. (1998) study, in that there was no TBI group for comparison purposes.

Strauss et al. (2001) used latency scores on the VSVT and CDCT (computer dot counting test) to predict under performance. These two malingering tests were administered within a neuropsychological battery of tests. Their study looked at performance differences between a naïve group, an experienced group who had professional contact with head injured patients and a TBI (no litigation) group that varied in severity but consisted of mainly mild injury. For each of these three groups, participants were randomly assigned to either a control or malingering condition.

malingering groups were given a scenario and asked to imitate a realistic head injury impairment. The computerized version of the dot counting test was similar to that of the paper and pencil version; however, the computerized version required the participant to enter the number of dots he/she had seen on a computer keyboard. If the answer was incorrect, the question remained on the computer until the correct answer was given. A potential disadvantage to this method is asking the examinee to enter his/her answer may increase variability due to individual differences as opposed to having the examiner enter the answer and a tone to provide feedback.

Their results showed that the RT for the VSVT was more variable for the malingering group; more so for the hard items than for the easy ones. The control group obtained shorter RTs than the malingering group, and shorter RTs were found for easy items than for hard items. For the CDCT, the malingering group took longer. In addition, ungrouped dots produced a longer response time rather than did the grouped dots. Classification rates for intraindividual variability found the RTs for the VSVT predicted 85% of the controls and 59% of the malingerers. The CDCT-ungrouped condition provided better classification than the grouped, with 87.7% accuracy for controls and 64.7% for malingerers.

Only one study found that RT scores may not be helpful in detecting malingering. Bolan, Foster, Schmand, & Bolan (2002) compared latency scores on three different malingering tests: the DMT, the Amsterdam Short Term Memory Test (ASTM; Schmand, de Sterke, & Lindeboom, 1999), and the TOMM-C. However, only the latter two tests were mentioned in their results. Twenty undergraduate students were randomly assigned to a control group or simulator group. The simulator group had

longer latencies than the control group on both tests. In this regard, these findings support those of Rees et al. (1998). However, a small overlap occurred on both the ASTM and the TOMM-C for the control and simulator group. Thus, cut-offs to differentiate groups could not be established. Bolan et al. (2002) therefore concluded that response latency information on its own was insufficient in detecting malingering. The methodology in this study has several potential problems. First, requiring the participants to type in their own responses on the ASTM adds individual variability. Additionally, the tests were not embedded within other neuropsychological tests to disguise the purpose of the study and finally no TBI group was used.

The Paced Auditory Serial Addition Test (PASAT) and Malingering

Strauss, Spellacy, Hunter, & Berry (1994) studied the PASAT and a simple reaction test (SRT) as a potential way of detecting malingering. They used patients with mild to moderate head injuries (n=28), healthy control simulators (n=9), and a control group consisting of undergraduate students (n=10). Results showed that for both the SRT and the PASAT, malingerers had the longest RT, with the head injury group next and controls performing the fastest. The SRT yielded a prediction rate of 75.5% based on reaction time scores. Unfortunately, the PASAT, proved to be a poor tool in that it showed low sensitivity and specificity. One reason behind the poor classification rating of the PASAT may have been that only two intervals (2.0 and 1.6), as opposed to the original four intervals, were used. In addition, the head injury group was not matched to the education level of the control participants and some of these individuals were involved in litigation, which in turn may have increased the likelihood of malingering.

Reaction Time Tests and Malingering

In addition to using latency scores on conventional malingering test as a way of detecting malingering, one may use RTs generated from reaction time tasks.

Unfortunately, only one study has investigated using reaction time tasks as a way of detecting under performance. Wogar, Broek, Bradshaw, & Szabadi (1998), using 20 control participants, 20 simulators, and 25 TBI patients (post-injury greater than 12 hours), examined each group's performance in a matching-to-sample task. It is not clear as to the severity of the head injury group used in the Wogar et al. study. The matching-to-sample task was embedded within other neuropsychological tests. Participants completed eight matching-to-sample tasks, each varying in difficulty. Each trial started with a 100ms beep and a stimulus presented at the top center of the computer screen. The stimulus consisted of a row of letters that varied in number. Two comparison stimuli were shown at the bottom left and right screen, one identical to the sample stimulus the other differed by one letter. No delay interval between the sample presentation and comparison stimuli was noted in their study.

The study found a decreased linear relationship between RT and the number of letters presented. Moreover, the TBI group had a steeper slope than either the control or simulator group. The control and simulator groups had similar slopes but the simulator group had a longer RTs. They concluded that simulators were unable to mimic performance of TBI patients but that latency scores could potentially be useful in detecting malingering.

Summary

Any attempt to compare the results of the RT studies reviewed above is faced with a variety of procedural inconsistencies that make it virtually impossible to arrive at any conclusion as to the role that RT measures have in detecting malingering. First, many studies failed to provide a comparison TBI group. This group is imperative for we wish to ensure the latency scores associated with the tests are relatively insensitive to the impairment that the malingerer is attempting to simulate. Second, the tests used to detect malingering were not embedded within other tests as to ensure the participant is not “tipped off” to the purpose of the study. Third, participants were not matched to ensure there were no age effects. Fourth, no questionnaires examined such variable as how successful the participants thought they were in faking the impairment, and what strategies they used. Fifth, in many cases the RT results were not compared to those produced by other tests that were designed specifically to detect malingering.

As reviewed, past studies have shown improvements in the area of detecting malingering. However, as information regarding these tests becomes more readily available to the general public, attempts to malingering may go undetected. Thus, we require new and improved ways of detecting malingering or amending current ones.

Given that reaction time tasks have proven to be an effective and sensitive tool in indexing levels of head injury, an attempt to employ these tests further as a way of detecting malingering may prove useful. An added benefit in using a reaction time task as a way of looking at malingering is that it may provide the examiner with a more sensitive tool in detecting malingering. That is, in addition to measuring the correctness

of a response, reaction time tests may provide an index on the amount of time that is required to plan and execute the malingered response.

However, few studies have attempted to evaluate the possibility that reaction times provide a sensitive measure to detect malingering. Although latency scores obtained from various types of psychological tests have been used to identify intentional under-performance, only one has used reaction time tests per se as a potential method of detection. Consequently, the purpose of the current study was to determine if reaction time tests were sensitive to malingering and, if so, how did they compare to other malingering tests.

To ensure a successful detection rate for malingering, the test has to be insensitive to the type of impairment that the person was attempting to feign (e.g., memory impairment in TBI). Therefore, it was important to compare the performance of a Malingering Group to that of a TBI Group, as well as to that of a Cognitively Intact Group. Additionally, comparing the differences between malingering tests was also of interest: Did all three malingering tests provide similar results? Which test was the best?

It was predicted that reaction time differences would result between the three groups. It was also speculated that the TBI Group's performance would progressively decline as each CTIP test became more complex. That is, latency scores would get progressively worse, more so than that of the Control Group. It was hypothesized that the greatest difference in reaction time would be found on the simple reaction task measure, with the TBI and Control Group performing somewhat similarly on this task and the Malingering Group performing well above (i.e., increased RT) the other two groups.

Finally, it was anticipated that the greatest variability would occur within the Malingering group. That is, not only would these participants tend to overestimate how slowly a TBI patient would perform, the decision processes involved in formulating and executing their malingering strategy would increase the variability of their performance.

METHOD

Participants

The study involved four groups: Malingers, Controls, Mild TBI and Moderate TBI. All groups originally had 15 participants (one individual from the Control group and one individual from the Malingering group were dropped due to unusual test scores, that is, the participants produced long reaction time responses that resulted in outlier scores). The age for the university student population (Control and Malingering participants) ranged from 18 to 24 years ($M = 19.8$, $SD = 1.47$); for the Mild TBI population the age ranged from 18 to 25 years ($M = 21$, $SD = 1.81$) and for the Moderate TBI group age ranged from 19 to 59 years ($M = 44.8$, $SD = 12.59$). Gender distribution for each group was as follows: Control (9 males, 4 females), Malingering (4 males, 11 females), Mild TBI (11 males, 4 females), Moderate TBI (8 males, 7 females). All participants had English as a first language. The Malingering and Control Groups were students recruited from Carleton University who received course credit for their participation. The TBI patients were recruited from an ongoing study that was directed by Dr. Marshall at The Rehabilitation Center in Ottawa. The patients were paid \$50 by Dr. Marshall to participate in his study. Any participant from the TBI group that was involved in litigation was dropped from the study.

Procedure

Students were randomly assigned to either the Control or Malingering Group.

The Malingering Group was given the following set of instructions over the phone:

Scenario

In this study you will be asked to complete a set of tasks that are often used to measure a variety of changes that occur in people who have brain damage. As you take each test, we would like you to assume the role of someone who has experienced some brain damage from a car accident.

Pretend that you are involved in a head-on collision. You hit your head against the windshield and were unconscious for 15 minutes. You were hospitalized overnight for observation and then released. Gradually, over the past few months, you have started to feel normal again. However, your lawyer has informed you that you may get a larger settlement from the court if you look like you are still suffering from brain damage. In the real world, the usual purpose of the tests you are about to take is to determine if the accident has produced any impairment in your abilities due to brain damage.

As you portray the above person, try to approach each test as you imagine this person would respond if he or she had been given the same instructions from his or her lawyer or someone else hoping to influence the amount of the settlement. Keep in mind that settlement monies depend upon your being diagnosed as cognitively impaired on these tests. Also be aware that having a lawsuit pending often raises the suspicion that people may try to exaggerate their difficulties. That means your impairments resulting from the head injury must be believable. Major exaggerations, such as not being able to do anything, remembering absolutely nothing, or completely failing to respond, are easy to detect.

The Malingering Group was given one week to prepare and were told they could use any resources to do so. When these participants returned one week later the scenario was reread. The Control and TBI Groups received the same procedure except that they did not read the scenario and instead were asked to perform to the best of their ability.

Demographics for all participants were obtained prior to testing, as was a medical history for the TBI Group. Since the TBI groups received a different neuropsychological battery as part of another study, although performance on the CTIP for all four groups was examined, only the Malingering Group and Control Group were compared on the conventional malingering tests. Furthermore, all participants read and signed an informed consent form. Testing took approximately 2 hours.

Test administration for the Malingering group and the Control group occurred in the following order: Trials A and B, Digit Span, TOMM-C (learning trials), CTIP, Digit Symbol, TOMM-C (retention trials), 20-Item Word List, VSVT, Adjusting-PVSAT, follow-up questionnaire, and debriefing.

Materials

Computerized Test of Information Processing (CTIP; Tombaugh and Rees, 1999).

The CTIP was divided into the following three components. All reaction times and number correct/incorrect were recorded via the computer program. Each task had 50 trials for the Malingering and Control group; while the TBI groups were administered 30 trials.

Simple Reaction Time: The examinee pressed the space bar, with their dominant hand, each time an 'X' appears in the middle of the computer screen.

Choice Reaction Time: The participant pressed a right key every time the word 'duck' appears, and the left key when the word 'kite' appeared.

Semantic Reaction Time: The participant had to decide if a word belonged to a specific category. On each trial one of four categories (Weapon, Furniture, Bird, Fruit) was randomly presented on the screen. Approximately 2 sec later a word

appeared below that category that may or may not have been a member of that category. If the word belonged to the category the right key was depressed. If the word did not belong to the category the left key was pressed.

Malingering Tests

Test of Memory Malingering-Computerized (TOMM-C; Tombaugh, 1996): This test required the examinee to study 50 pictures for 3 seconds each. Immediately afterward, a two forced-choice recognition procedure was employed (retention trial) whereby 1 old and 1 new picture was shown. The participant chose the picture they saw previously and feedback of correctness was provided. This administration was done twice. After the second administration a 15 min delay occurred and a final 2-choice recognition test was then given.

Victoria Symptom Validity Test (VSVT: Slick et al., 1997): The VSVT involved the presentation of a string of 5-digit numbers (e.g., '34098') via a computer. A forced-choice recognition task was then given whereby the previous seen string of digits was paired with a new 5-digit number (foils). The examinee was required to select the number previously shown as quickly as possible. Foils consisted of either 24 easy items or 24 hard items. Easy items were foils that share no common digits with the original string shown, for instance '34028' vs. the foil '56197'. Hard items were considered foil items that invert or transverse the middle two digits while the remaining digits stay in the same sequence, for example '46923' vs. '46293'. There were a total of 48 trials in three blocks of 16 items each. Blocks had increasing interval times between digit presentation and

recognition trials from 5s in the first block to 10s in the second and 15s in the third.

20-Item Word List: A word list containing 20 nouns was visually presented to the examinee and simultaneously read out loud by the experimenter. Each word was presented on the computer for 1 second. A two-alternative forced-choice task was given whereby the participant chose between the target word and a foil. This test was a modified version of the 21-Item Word List (Iverson et al., 1991) that had been adapted for computerized presentation.

Neuropsychological Tests of Attention

Digit Span (Subtest from the LAMB: Schmidt & Tombaugh, 1995): Participants repeated a series of progressively longer digits in the same order for the first part, and in the reverse order for the second. Two trials were given for each number series. Digit span length was the number of digits that were recalled in the correct sequence for the longest series. Each correctly remembered series was given one point and the total number of points was recorded for both the forward and backward span.

Trail Making Test A and B (Reitan & Wolfson, 1985): On part A, the participant joined 25 numbers in order as rapidly as possible. On part B, 13 numbers and 12 letters were joined in order starting with a number first, then a letter, etc in the same manner as for part A. The score for each part was the time to complete each task.

Symbol Digit Modalities (SDMT, Smith, 1982): Participants were required to pair numbers associated with 9 particular symbols according to a key. Ninety seconds

were given to complete this task and the resulting score was the number of symbols correctly identified.

Adjusting-Paced Visual Serial Addition Task (Adjusting-PVSAT: Tombaugh, 1999): The A-PVSAT visually presented a series of digits from 1 to 9. A mask to prevent after-image formation followed each of the visually presented digits. Each administration began with a practice session of 12 trials. The ISI length between each digit remained constant at 3.0s. Following the practice session, a series of 200 trials began with the presentation of a single-digit number followed by a 2.4s interval. Following this interval, the next digit was presented and was added to the previously presented digit. The participant said the answer aloud and the experimenter typed the response into the computer using the numeric keypad. The ISI was adjusted with a titrated/stairstep procedure so that each correct response was followed by a 20 ms ISI decrease and each incorrect response was followed by a 20 ms increase. This procedure resulted in a 50% threshold.

Questionnaire

Appendix A shows a series of questions that were posed to the Malingering Group to determine how genuine and successful their attempts to malingering were, as well as strategies they implemented.

RESULTS

Computer Test of Information Processing

CTIP: A primary purpose of this study was to determine the usefulness of a new detection method, the CTIP, in identifying individuals attempting to malingering a brain injury. As previously mentioned, the CTIP uses three different reaction time tasks, which

for the purpose of the following section we will call levels of ‘complexity’ (Simple, Choice, and Semantic).

The CTIP results are shown in Figure 1. A repeated measures analysis of variance (ANOVA) performed on the data revealed the following: Group: $F(3, 55)=12.11, p<.001$; Complexity: $F(2, 110)=62.10, p<.001$; Group x Complexity: $F(6, 110)=3.49, p<.003$.

Figure 1. Reaction time (ms) on the CTIP for Malingerers, Controls, Mild TBI and Moderate TBI Groups.

Inspection of Figure 1, in conjunction with a LSD pair wise comparisons, reveals that the group x complexity interaction was due to several factors. First, the performance of the Control and Mild TBI Groups was virtually identical at all levels of complexity. No pair wise comparisons between these two groups were significant for any of the three CTIP tests. Thus, it appears that the CTIP does not differentiate between these two groups. Table 1 detailed the group means and confidence intervals for each CTIP test.

Table 1. Means and Confidence Intervals on each CTIP test for Malingering, Control, Mild TBI, and Moderate TBI Groups

Simple CTIP test			
Group	Mean	95% Confidence Interval	
		Lower	Upper
Malingering	884.73	750.02	1019.45
Control	286.82	147.38	426.27
Mild TBI	293.60	158.88	428.32
Moderate TBI	429.57	349.94	1043.53
Choice CTIP test			
Group	Mean	95% Confidence Interval	
		Lower	Upper
Malingering	1130.73	909.61	1351.86
Control	498.82	269.94	737.71
Mild TBI	514.93	293.81	736.06
Moderate TBI	799.79	570.90	1028.67
Semantic CTIP test			
Group	Mean	95% Confidence Interval	
		Lower	Upper
Malingering	1781.63	1434.84	2128.43
Control	680.96	322.00	1039.93
Mild TBI	696.73	349.94	1043.53
Moderate TBI	1339.64	980.68	1696.61

Second, the reaction time scores for the Malingering Group were significantly higher than those for the Control and Mild TBI Groups. The semantic search produces the greatest differential effect between the Malingering group and the Control/Mild TBI Groups. The obvious implication of these results is that semantic search should be better able to detect malingering in mild TBI patients than either the simple or the choice.

Third, increasing complexity caused the performance differences between the Moderate TBI Group and the Control/Mild TBI Groups to become increasingly larger. The greatest differences between the Malingering and Moderate TBI Groups occur for the simple test. As complexity increases, the Malingering and Moderate TBI Groups become slightly more similar. Thus, in contrast to the effects found with Mild TBI participants, the semantic search should be relatively less effective in detecting Malingering for individuals with Moderate TBI while simple RT procedures should be the most effective.

However, further inquiry seemed warranted as to what the latency data actually represented for the semantic level of complexity. That is, does a general latency slowing occur or do the “same” (e.g. furniture/chair) or “different” (e.g. furniture/lemon) responses produce different latencies? As shown in Table 2, for all groups the “same” categorical response was faster than the “different” categorical response. If a general slowing in latencies were to occur, percent change ratios should be minimal, while a differentiated response style would be represented by a substantial percent change between the “same” and “different” responses.

Table 2. Semantic categorical responses (ms) on the CTIP.

Group	<u>CTIP Semantic Category</u>	
	Same	Different
Malingering	1667	2104
Control	681	715
Mild	661	738
Moderate	1194	1456

To determine if differences occurred between responding to a same categorical stimulus than to a different categorical stimulus, a percent ratio of change was calculated, formula (Different-Same/Same) x 100. Figure 2 details the percent change between the “same” and “different” responses.

Figure 2. Percent change for same and different responses on semantic search.

Inspection of the graph indicates the smallest percent change occurred for the Control Group, while the largest change occurred for the Malingering Group. Again, the Control and Mild TBI scores were somewhat similar as were scores for the Malingering

Group and Moderate TBI Group. Significance between group differences are shown in Table 3 following a one-way ANOVA.

Significant differences in percent ratio change were found between the Malingering Group and the Control Group as well as between the Malingering Group and the Mild TBI Group. As previously noted, all groups tended to differentiate their response as opposed to supplying a general latency constant. Since significant differences were found between the Malingering Group and both the Control and Mild TBI groups, it appears that the Malingering Group purposefully responded differently than did the other two groups.

Table 3. Multiple comparison results for percent ratio change on CTIP semantic task.

Group Comparisons	Significant
Malinger vs. Control	Yes
Malinger vs. Mild TBI	Yes
Malinger vs. Moderate TBI	No
Control vs. Mild TBI	No
Control vs. Moderate TBI	No
Mild TBI vs. Moderate TBI	No

Yes =significant; No=not significant

While the above analyses clearly indicate that reaction time scores are different for malingerers and non-malingerers, they do not provide any clear cut criteria for classifying whether a person is malingering. Consequently, the frequency distribution scores were examined on each of the CTIP tests to determine optimal cut-off scores. Cut-

off scores were determined by establishing values that would result in the highest specificity rating for the Control Group, as more concern was given to misclassifying an individual as a malingerer than allowing an individual who is actually malingering to go undetected. Table 4 shows the cut-off scores for each of the CTIP tasks and sensitivity/specificity ratings for these respective cut-off scores.

Table 4. Sensitivity/Specificity ratings for CTIP latency scores

CTIP Test	Cut-score	Sensitivity	Specificity		
			Control	Mild TBI	Moderate TBI
Simple	>520	80%(12/15)	100%(14/14)	93%(14/15)	93%(14/15)
Choice	>650	80%(12/15)	100%(14/14)	93%(14/15)	73%(11/15)
Semantic	>1140	80%(12/15)	100%(14/14)	93%(14/15)	47%(7/15)

For example, on the simple RT task, when a cut-off score of greater than 520ms is used a 80% sensitivity (Malingering), and a 100% (Control), 93% (Mild), and 93% (Moderate) specificity ratings resulted. Put another way, 12 out of 15 participants in the Malingering Group were correctly identified as a malingerer. For the Control Group, 14 out of 14 participants were classified correctly as not being malingerers, while in both the Mild TBI and Moderate TBI Group, 14 out of 15 of the participants were correctly identified as not being malingerers.

Inspection of Table 4 indicates that as the complexity of the CTIP tasks increase the sensitivity and specificity ratings for the Control and Mild TBI Groups remain the same; however, the specificity for the Moderate TBI Group decreases. This result is not

surprising, as individuals with more impairment will have greater difficulty as tasks become more challenging.

Variability

Because variability has been known to increase with TBI individuals (Stuss et al., 1994; Van Zomeren and Brouwer, 1989), the CTIP scores were examined to determine if similarities or differences in variability occurred between groups. If individuals in the Malingering Group, in fact, overestimated their performance, it was likely that variability for this group would also be exaggerated.

There are three common ways to measure variability: standard deviation, coefficient of variation, and semi-interquartile range. The most commonly used measure is standard deviation. Using this method, the current study found group differences in variability. However a common criticism of this method is that it does not take into account the fact that greater variability is normally associated with higher mean performance. Thus, a different way to measure variability was employed: coefficient of variation. The coefficient of variation examines each standard deviation relative to the mean. In this regard, larger mean differences are accounted for when using this type of measurement. However, coefficient of variation can be criticized if the data is skewed whereby larger means can affect the outcome of the analysis.

Van Zomeren and Brouwer (1989) and Stuss (1994) used the above measures to examine variability differences between healthy individuals and those with a TBI. Both studies found significant group differences in variability, with the TBI Group obtaining greater variability than the Control Group. The current study replicated these results. However, as mentioned, these measures of variability are not without criticism. Extreme

scores can influence both standard deviation and coefficient of variation measures of variability. Thus, a semi-interquartile measure ($(25^{\text{th}} Q - 75^{\text{th}} Q)/2$) was used as it is relatively insensitive to data that is skewed, which occurred for some groups in the present study.

A semi-interquartile range provides a relative measure of dispersion by eliminating the quadrants where most of the skewness occurs. Variability was assessed using both 50 trials (Control and Malingering Groups) and 30 trials (Mild TBI and Moderate TBI Groups). For the TBI group only 30 trials were used due to a difference in testing procedures. However, since no differences between the 50 trials and 30 trials occurred for the Control and Malingering group, analysis was done over 30 trials so as to maintain continuity between all groups (see Table 5).

Table 5. Semi-interquartile range for CTIP tests

Group	CTIP Test		
	Simple	Choice	Semantic
Maligner	238.26 (189.26)	382.76 (308.71)	610.83 (414.74)
Moderate TBI	56.90 (59.35)	153.64 (141.67)	276.23 (157.70)
Mild TBI	29.46 (21.92)	80.73 (36.47)	138.76 (95.55)
Control	24.67 (10.69)	83.96 (36.83)	109.53 (39.24)
Semi-interquartile range = $(25^{\text{th}} Q - 75^{\text{th}} Q)/2$ (SD)			

By way of a one-way ANOVA, the variability for each group increased as the task became harder. Additionally, the variability for the Malingering Group was greater than all other groups for each level of complexity SRT: $F(3,58)=15.12$, $p<.001$, Choice:

$F(3,58)=9.96, p<.001$, Semantic: $F(3,58)=14.81, p<.001$. By way of pair-wise comparisons, the Malingering Group obtained significantly greater semi-interquartile variability than all other groups.

Van Zomeren and Brouwer (1989) asserted that when examining reaction time and variability for TBI patients and control participants the two variables were proportional to one another. In other words, the reaction time is highly correlated to the variance for both groups, and as reaction time increases variability also increases. To determine if the current study's results supported those of Van Zomeren and Brouwer, a ratio of semi-interquartile values to median values were evaluated. Also of interest was to examine if the Malingering Group would obtain different or similar ratios. Table 6 shows these ratios for each group at various levels of complexity.

Table 6: Ratio of semi-interquartile to median scores for each group on the CTIP tests.

Group	CTIP Test		
	Simple	Choice	Semantic
Malingering	0.28	0.34	0.33
Control	0.09	0.17	0.16
Mild TBI	0.10	0.16	0.19
Moderate TBI	0.10	0.19	0.20

Ratio similarities between the Control Group and Mild TBI Group as well as between the Control Group and Moderate TBI Group supports previous findings of Van Zomeren and Brouwer (1989). However, the ratio for the Malingering Group, as found by a one-way ANOVA, significantly differed from all other groups and for each CTIP

test, SRT: $F(3, 58)=6.67, p<.001$, Choice: $F(3, 58)=5.80, p<.002$, Semantic: $F(3, 58)=9.62, p<.001$. In other words, for the Malingering Group more variance relative to the median was found. This result is not evident for any other group. Perhaps this ratio difference can be used as an indication of when an individual malingers.

A semi-interquartile measure provides a relatively clear picture of the variability with the extreme scores eliminated, but what about examining just these outlining scores? Using a percentile ranking system allowed for variability to be examined at the extreme ends of the distribution, as well as at intermediate levels. Figures 3,4, and 5 show percentile ranks for each group at each level of complexity. Notice that the mean reaction time is greater for the Malingering Group for each percentile rank level. This result occurs for each level of complexity. Of particular importance is that at the mean reaction times at the 10 percentile, which reflects the scores having the longest latencies (i.e., outliers), are dramatically longer for the Malingering group. That is, the Malingering group performs the worst at this level as indicated by the longer reaction time scores. Therefore, the Malingering group at the 10th percentile can be easily distinguished from all other groups.

A one-way ANOVA verified this effect: SRT: $F(3, 58)=12.07, p<.001$, Choice: $F(3, 58)=7.35, p<.001$, Semantic: $F(3, 58)=7.60, p<.001$. Further examination by way of a pair wise comparison found significant differences between the Malingering Group and the Moderate TBI Group at the 10th percentile rank for each level of complexity. That is, although the mean reaction time for the Malingering Group and the Moderate TBI Group became more similar as complexity increased, the differences between the two groups remained significant.

Figure 3: Percentile Rank for simple CTIP¹

Figure 4: Percentile Rank for choice CTIP

Figure 5: Percentile Rank for semantic CTIP

¹ Examination of the 10th percentile rank is similar to an Ex-Gaussian procedure (Leth-Steensen et al. 2002).

Conventional Malingering Tests

Another purpose of the study was to compare the ability of three conventional tests used to detect malingering and to determine if latency scores on these tests are sensitive to malingering. Before comparing latency scores results, inspection of accuracy using the number of correct responses was undertaken. Again, due to the lack of malingering tests for the clinical data, that is, the TBI groups received a different neuropsychological battery as part of another study, only the Malingering Group and Control Group were compared.

Test of Memory Malingering (TOMM-C): As anticipated, a repeated measures analysis found a significant group effect for number of correct responses, $F(1,27)=19.55$, $p<.001$. Figure 6 clearly indicates these group differences for trial 1 (T1), trial 2 (T2) and retention (T3).

Figure 6. Number of correct responses for each trial on the TOMM-C for the Malingering and Control Groups.

The next step was to look at overall latency scores for each trial of the TOMM. Here, both a significant group effect and a trial effect resulted: $F(1,27)=23.91$, $p<.001$; $F(2, 54)=22.41$, $p<.001$, respectively. Figure 7 shows these results. Since overall

latencies were examined, the question of whether correct latencies would produce similar effects to those seen for overall latencies (which include both correct and incorrect responses) remained. Figure 8 shows latency scores that are based solely on correct responses. Results based upon Figure 8 produced a significant group effect and a trial effect, $F(1,27)=19.74$, $p<.001$; $F(2,54)=7.31$, $p<.002$, respectively. Thus, the answer to the question was, in fact, yes. Results showed similarities between overall latency scores and correct latency scores.

Figure 7. Response latencies for all responses on the TOMM-C for the Malingering and Control Groups.

Figure 8. Response latencies for correct and incorrect responses on the TOMM-C for the Malingering and Control Groups.

One question remained, would individuals in the Malingering Group obtain similar latencies for correct and incorrect responses? It should be noted that incorrect responses for the Control Group were not examined since too few incorrect answers were obtained. Figure 8 shows the differences between latency scores for the Malingering Group on correct and incorrect responses. By way of a paired t-test no significant results for across trials $t(11)=-.354$, $p=.730$ were found between correct and incorrect latency scores for each trial. Therefore, on the TOMM, individuals in the Malingering Group produced similar response styles for correct and incorrect responses.

20-Item Word Test: Table 7 summarizes the differences between groups on the 20-Item Word Test. Since there was only one trial involved in the 20-item Word Test, a one-way ANOVA was used to analyze differences between groups. Again, number of correct responses were examined first and clear differences resulted between groups, $F(1,26)=15.3$, $p<.001$. A similar process to what was done when examining the TOMM was employed. Overall latency scores were examined and then correct latency scores, both producing similar results, $F(1,27)=9.06$, $p<.006$; $F(1,27)=7.46$, $p<.011$, respectively. Again, the question of whether differences in the Malingering Group occurred between correct and incorrect responses remained. Examining these differences using a paired t-test showed no significant results, $t(13)=-.554$, $p=.589$. Hence, as with the TOMM, it looks as though individuals attempting to malingering on the 20-Item Word Test do so generally speaking, meaning their exaggerated time to react is the same across all spectrums.

Table 7. Number of correct responses and latencies (ms) for the 20-Item Word Test.

	Malingering	Control
# Correct Responses	12.9	17.4
Latency (ms)		
Overall responses	2046	1151
Correct responses	2014	1233
Incorrect response	2083	

Victoria Symptom Validity Test (VSVT): As with the other Malingering tests, the VSVT detects malingering on the basis of the number of correct responses. In addition, it assumes that an individual attempting to malingering will decrease the number of correct responses as the perceived difficulty increases; that is, as the time between the study and recognition trial becomes longer. The mean number of correct responses on each trial for the two groups is shown in Figure 9. Repeated measures ANOVAs substantiated the observation that the Control Group consistently had a greater number of correct responses than did the Malingering Group, Group: $F(1,26)=26.04$, $p<.001$; Trial: $F(1,26)=1.40$, $p=0.257$; Group x Trial: $F(1,26)=1.84$, $p=0.17$. The failure to find a significant trial effect or group x trial interaction fails to support the notion that as intervals increase the task becomes more difficult (or it is perceived to be more difficult).

In order to determine if the “easy” and “hard” number sequences exerted any differential effect on responding, performance was evaluated over the easy/hard variable (see Figure 10). When examining the easy/hard variable for number of correct responses on each trial, a 2x2x3 factorial analyses showed a group effect ($F(1,26)=23.19$, $p<.001$)

an easy/hard effect ($F(1,26)=18.20, p<.001$), and a group x easy/hard interaction ($F(1,26)=7.06, p<.013$). As previously mentioned no trials effect occurred (i.e. no differences across 5sec, 10sec and 15sec trials), therefore this analysis is similar to a 2x2 analysis that collapses over trials.

Figure 9. Number of correct responses on each trial for the Malingering and Control Groups on the VSVT.

Figure 10. Number of correct response for easy and hard number on each trial for the Malingering and Control Groups on the VSVT.

Further exploration into latency scores yielded results similar to those seen with the other two malingering tests (see Figure 11). A repeated measures ANOVA showed that the Malingering Group had longer latencies than the Control Group and the latencies progressively increased over the three trials: Group: $F(1,26)=33.48, p<.001$, Trials: $F(2,52)=4.20, p<.02$. Thus, the results are similar to those seen for number of responses between the groups, except that as difficulty increases a progressive effect occurs in that all individuals take longer to respond.

Figure 11. Overall latency scores for each trial on the VSVT.

Since the overall latency data reflect the combined effect of both the easy and hard number sequences, further analyses were conducted to see the relative contribution of each to the overall latency score. Figure 12 shows the easy and hard latencies scores for each group. A 2x2x3 analyses revealed a group effect ($F(1,26)=23.7, p<.001$), an easy/hard effect ($F(1,26)=24.97, p<.001$), and a trials effect ($F(2,52)=11.26, p<.001$).

Figure 12. Easy and Hard latencies for Malingering and Control Group on the VSVT.

A final series of analyses were undertaken to determine if different latency occurred on correct and incorrect response. The data used in these analyses are shown in Figure 13. The first ANOVA examined whether differences existed between the two groups when correct responses were emitted. The latencies for the Malingering Group were significantly longer than for the Control group and progressively increased as the time intervals increased: Group: $F(1,26)=20.90$, $p<.001$; Trial: $F(2,52)=5.51$, $p<.007$.

Figure 13. Correct/Incorrect Latencies for Malingering and Control Group for each trial on the VSVT.

A second analysis was undertaken to determine if a difference existed between the correct and incorrect latency scores for only the Malingering Group (the small number of incorrect responses from the Control Group prevented a similar analysis to be preformed for this group). A paired t-test analysis showed that incorrect responses took longer than correct responses ($t(12)=-3.29, p<.007$).

Sensitivity and Specificity: The differences between the three conventional malingering tests were examined to see if they equally support detecting individuals attempting to malingering or if one test is superior to another. Table 8 indicates the cut-off criteria (taken from the tests respective manuals) for each test and provides sensitivity/specificity findings.

Indications were that the TOMM might be preferable in determining group classification as the test is known to have high validity and is widely used. Results provided confirmation of this as the TOMM achieved the highest sensitivity (73%) and specificity (100%) ratings. That is, 73% of the participants were correctly classified as malingeringers, while a 100% of the participants in the control group were identified as not malingering.

For the 20-Item Word Test, a cut-off score at the 80th percentile was preferable to that of the 90th percentile as fewer of the control participants were mislabeled as malingeringers. Here, 71.4% of the participants were correctly classified into the Malingering Group and 78.5% of individuals in the Control group were correctly identified as such. A cut-off score of less than 14 was used to minimize false positives, that is, 100% of the Control Group was correctly classified. Implementation of this cut-off scores resulted in a 57% sensitivity rating.

Although still a low rating, the VSVT achieved a higher classification rating for hard items than for easy items. Using hard items to predict classification, a total of 57% of the individuals were correctly classified into the Malingering Group and 100% classification was achieved for the Control Group. Reasons as to why these findings are lower than what previous studies have indicated will be elaborated on in the discussion.

Table 8. Specificity/sensitivity ratings for conventional malingering tests.

Test	Criteria	Sensitivity	Specificity
<u>TOMM</u>	<45 on T2 or ret. OR <chance on any trial	73% (11/15)	100% (14/14)
<u>20-Item</u>			
80%	<16/20	71.4% (10/14)	78.5% (11/14)
90%	<18/20	92.8% (13/14)	42.8% (6/14)
Min. False Positives	<14/20	57.1% (8/14)	100% (14/14)
<u>VSVT</u>			
Overall (n=48)			
Invalid	17-0	7% (1/14)	100% (14/14)
Questionable	29-18	29% (4/14)	100% (14/14)
Total	29-0	36% (5/14)	100% (14/14)
Easy (n=24)			
Invalid	7-0	7% (1/14)	100% (14/14)
Questionable	15-8	15% (2/14)	100% (14/14)
Total	15-0	21% (3/14)	100% (14/14)
Hard (n=24)			
Invalid	7-0	15% (2/14)	100% (14/14)
Questionable	15-8	43% (6/14)	100% (14/14)
Total	15-0	57% (8/14)	100% (14/14)

Strategies

Questionnaires were evaluated to determine the following for individuals in the Malingering group: how they prepared, strategies they employed, and how successful they thought they were. In relation to a scale of 1 to 5 (1=didn't try at all, 5=tried very hard; see Appendix) to determine how hard they tried to follow instructions ($M = 4.00$, $SD = 0.73$). Results indicating how successful participants felt they were in following instructions found ($M = 3.5$, $SD = 1.02$). With the exception of 3 participants in the Malingering group who indicated they did not prepare for the testing session, many of the remaining participants indicated they either used books, TV, course notes and/or the Internet.

The most popular strategy employed was to slow down reaction time, 8 participants indicated they used this strategy while the remaining participants mentioned such strategies as being confused or faking memory impairment. When answering which symptom(s) they felt would most likely be exhibited when attempting to fake a brain injury, 8 participants circled memory impairment, 5 circled attention, 1 circled both and 1 participant's questionnaire was incomplete.

DISCUSSION

The current study had two main purposes. One purpose was to look at the CTIP, a pure RT task, as a new method to detect malingering. The other purpose was to compare three conventional detection methods and to examine if latency scores could be used to detect malingering. The following discussion contains a detailed explanation of results, clinical implications, and future research related to each purpose.

Computer Test of Information Processing (CTIP)

Reaction Time

A primary interest of the study was to examine how individuals attempting to malingering would perform on the CTIP and if the test, in turn, could be used as a viable way in detecting malingering. However, before discussing the results as they pertain to malingering it should be noted that the CTIP tasks did not differentiate between the Control Group and the Mild TBI Group. As previously mentioned, the original use of the CTIP was to look at differences between individuals with head injuries. The failure of the CTIP to differentiate between the Mild TBI and Control Groups may have occurred because the CTIP is only able to measure deficits in information processing for those who sustain more severe injuries. Impairment in information processing, as measured by reaction time tasks, is affected in individuals who sustain a brain injury (Alexander, 1995; van Zomeren & Brouwer, 1989). In addition, it seems that a positive correlation exists between reaction time and severity of head injury (Rose et al., 1995; Stuss et al., 1989; van Zomeren & Brouwer, 1989). That is, individuals with mild head injuries will indeed produce faster response times compared to those individuals who have sustained more severe injuries. The CTIP tasks are quite simplistic; it may be that even the semantic task, which is supposed to be the most challenging test, is too easy. In this case, the CTIP tasks may not detect slight deficits in information processing that occur in mild TBI individuals.

TBI Groups: Previous studies using pure reaction time tasks have examined differences between healthy Control participants, TBI patients, and individuals attempting to malingering. A study by Wogar et al. (1998) used a reaction time test (see

introduction for further details) to inspect differences between these groups. Their findings were similar to those of present study in that they found longer reaction times for the malingering group compared to the control group but the TBI group produced the longest latencies when compared to all other groups. However, the severity of the head injury group used in the Wogar et al. study was not clear. If more severe TBI patients had been used the results would not be all that surprising.

The current results between Mild TBI patients and Moderate TBI patients differ from those of Collins and Long (1996). In the present study, on both the simple reaction time and choice reaction time tasks, the Moderate TBI Group obtained slower latencies than the Mild TBI Group. In addition, the Mild TBI Group obtained similar reaction times to that of the Control Group. However, neither TBI group was significantly slower than the Control Group for both these tasks. Similarly, in the Collins and Long study, the TBI group consisted of two groups, Mild and Moderate however, unlike the present study their results found both TBI groups obtained slower reaction time than control participants. TBI group membership was based on the Halstead-Reitan Battery impairment index. No additional information as to how Collins and Long determine the differences between Mild and Moderate TBI was provided. However, their study found that for both their simple reaction time and choice reaction time tasks individuals with a TBI were significantly longer at responding than Control participants. Their simple reaction time task was similar to that used in the current study. Their choice reaction time task showed outlines of four circles presented horizontally. The participant was required to press the space bar when any two of the four circles were filled in. As with the CTIP, Collins and Long's reaction times tasks employed a buzzer to notify the

participant of an incorrect response. However, a warning signal was provided on their choice reaction time task indicating the stimuli was about to appear. This warning signal prepares the subject for an anticipatory response and thus reaction times would more than likely be faster. Unfortunately, no explanation can be provided as to why their study showed significant results in slowed reaction time for the Mild TBI Group and the current study did not.

Additionally, the Collins and Long study examined how motor responses may be a variable in increased reaction time. Their study controlled for this variable by holding motor response constant for each level of complexity. For both reaction time tasks only the space bar was pressed in response to a particular stimulus and not pressed in the absence of one. As mentioned, their study found an increase in reaction time for TBI patients. They asserted this increase was more representative of an attentional deficit as the motor response was held constant. The current study did not control for such motor responses as more than one key was used for the choice test and the semantic test. It may be that the increase in reaction times in conjunction with increased complexity may have been affected by additional motor demands. Therefore, the notion of motor response (i.e., right vs. left) should be further investigated.

Malingering Group: Results of the current study supported the prediction that individuals attempting to malingering would ‘overestimate’ the reaction time performance of someone with a head injury on the CTIP tasks. For all levels of complexity the Malingering Group tended to obtain slower latency scores than did all other groups, with the greatest misjudgment occurring at the semantic level. It seemed as though individuals in the Malingering Group were unable to correctly simulate bona fide TBI reaction times.

The inability to accurately simulate this cognitive behavior may be due, at least in part, to the extra thought processing that is required for malingering. That is, once an individual attempts to think about how to respond, he or she is that much slower than an individual who just naturally or automatically responds. Additionally, it may be that individuals attempting to malingering focus their attention on how many items they should get correct or incorrect, thereby unintentionally making reaction times slower. However, this reasoning only applies for the choice and semantic test, since the simple test requires only one response. Overall, these results are consistent with those of Wogar et al. (1998) who found longer latency times for the malingering group than for the control group but differs in that their TBI group obtained longer latency times than all other groups. As mentioned previously, it was not clear as to the severity of head injured patients used in the Wogar et al. study. If their study used mainly severe TBI patients these results are not all that surprising.

As complexity increases, the differences between the Malingering Group and the Mild TBI and Control Groups became more pronounced. However, the opposite effect occurred for the Moderate TBI group where the difference between this group and the Malingering Group became less pronounced. Intuitively, this result makes sense as individuals with a more severe head injury have increased difficulty with tasks that become more complex (that is, additional information processing is required). Since complex tasks are more sensitive to information processing than simple tasks, longer latencies occurred enabling one to better differentiate the Malingering Group from the Mild and Control Groups.

Which CTIP task is best?

It should be cautioned that conclusions from this study apply only to the situation where the three CTIP tests are administered in the order indicated. Since individuals were unaware of what lay ahead, starting with simple and moving to a more complex reaction time task probably worked in favor of developing a new detection method. That is, administering only one task or starting with a more complex task may not produce similar results. Future research may want to address the importance of administering the CTIP tests in a particular order.

It seems as though some CTIP tests are more useful in differentiating between certain groups than are other CTIP tests. The first question to address is: “Which CTIP test best differentiates the Malingering Group from the Mild TBI Group?” All CTIP tests make clear distinctions between these two groups; however, the largest difference occurred on the semantic task, thus this test does the best job. Since a large percentage of those involved in litigation are suspected of sustaining a mild head injury it makes sense to use the semantic test for detection. However, because the semantic task is sensitive to head injury, it may prove to be relatively less sensitive in distinguishing differences between individuals with a more severe head injury and individuals attempting to malingering a TBI. This possibility is not particularly bothersome because the sequelae of a mild head injury are more commonly malingered than those associated with a severe TBI because other detection methods (i.e., CT scan) cannot be employed to corroborate the cognitive effects of a mild TBI. Thus, since individuals are less likely to malingering moderate to severe head injuries, the current study was less interested in the results for the Moderate TBI Group.

A second question to address is: “Which CTIP test best distinguishes the Malingering Group from the Moderate TBI Group?” Results showed that the simple CTIP would be the best indicator in differentiating these two groups as the largest differences occurred at this level.

The final question is: “If only one CTIP test were to be administered, which test would best differentiate the Malingering Group from any TBI Group?” That is, if one were concerned at examining a malingering group to both a mild and moderate TBI population which task would be best? In answering this question, examination of which test resulted in the largest significant difference between the Malingering Group and both TBI Groups was required. Regardless of the severity of brain injury that a TBI patient sustains, the simple task would be preferable in making a clear distinction between these groups because the largest differences between both TBI Groups and the Malingering Group occurred on this task.

If the semantic task is preferable in one instance and the simple task is preferable for another instance, why then administer the choice task? Instead of administering all three CTIP tasks, why not administer just the simple and/or semantic portions, thereby reducing administration time? Eliminating the choice task would not be recommended at the present time since a full understanding of how participants perceive only one test is not known. That is, since administering the CTIP involves going from the simplest test to the most complex task, it is likely that participants perceive each subsequent test as more challenging. If only one test is administered there are no other tests for the participant to compare it to and thus different results may occur. Of course this statements hold true for all cases with the exception of the simple reaction time test as it

is the first test to be administered and no comparisons have yet to be made by the participant. As mentioned above, future research may want to closely examine the effect of administering the CTIP in different orders or administering only part of the CTIP.

Response Patterns for Same vs. Different Responses

Detailed reaction time findings for the semantic task were examined to determine if individuals attempting to malingering provided a generalized or differentiated response pattern. That is, did individuals in the Malingering Group vary their response style or respond consistently to “same” and “different” responses. For all groups “same” responses were faster than “different” responses. If the stimulus matches the category a response is made immediately, but if the stimulus word does not match the category an individual’s response is delayed until he or she continues to search the semantic lexicon and concludes that it’s not the same. To determine if differences occurred between responding to a same categorical stimulus than to a different categorical stimulus, a percent ratio of change was examined. Percent ratio changes were significantly higher for the Malingering Group compared to the Control and Mild TBI Groups. This difference indicates individuals attempting to malingering differentiate more between “same” and “different” responses than do individuals from the other groups. Possibly, since the semantic task involves more reasoning and thus more information processing, malingering individuals responded appropriately by slowing their answer more for a “different” response than for a “same” response. That is, although their latencies tend to be slower for both types of responses, they are proportionally slower for the “different” response. These results further support differences between an individual attempting to malingering and a mild TBI case. It seems that individuals attempting to malingering a head

injury attempt to manipulate how they react for “same” and “different” stimuli as opposed to merely responding consistently slower on all responses.

Variability Differences?

Because increased variability occurs in individuals with head injury (Van Zomeren and Brouwer, 1989), CTIP scores were examined to inspect similarities or differences between the TBI Groups and Malingering Group. Since individuals attempting to mangle overestimated the time it would take to respond, it also seemed reasonable to assume these individuals would have more variability in their scores. Variability was examined via several different methods. Using standard deviation and coefficient of variation the findings from the current study supported past findings (Stuss et al., 1994; Van Zomeren and Brouwer, 1989) with the TBI Group showing greater variability than the Control Group. In addition, the Malingering Group obtained the most pronounced variability of all groups. However, as already discussed in the results section, there are limitations to these methods as extreme scores can easily influence the outcome. Further variability differences were examined using semi-interquartile range, whereby extreme scores were eliminated. Results using this method again showed variability differences between groups, with more pronounced variability for the Malingering Group than for all other groups. Finally, percentile ratios were calculated in order to examine the tails of the distribution (i.e., extreme scores). Again, variability differences were found here, with the Malingering Group showing the greatest variability. In summary, all measures of variability showed that the Malingering Group obtained significantly greater variability in scores than any other group.

Van Zomeren and Brouwer asserted that an increase in variability is correlated with an increase in reaction time for TBI patients. Although, the coefficient of variation results supported this assertion, a ratio of semi-interquartile range was also examined. The difference between these two measures of variability is that the coefficient of variation is a ratio of the group standard deviation to the group mean, while the latter is a ratio of the semi-interquartile range to the group median. The latter ratio is less affected by extreme scores and therefore, may be a better control of the correlation between reaction time and variability.

As with the current study, Van Zomeren and Brouwer found the ratio of the semi-interquartile range to the median was similar for the Control and TBI Groups. However, it does not necessarily mean that examining variability scores provides evidence that an individual is malingering. Therefore, the current study went a step further in determining if similar ratios existed between individuals attempting to malingering and TBI patients. Sure enough, different ratios were found for the Malingering Group. This result is unique in that no other study has examined the differences in ratios between a Malingering Group and a TBI Group. This result nullifies the reasoning that increased variability is proportional to increased reaction time, at least for the Malingering Group. It may be possible to use this ratio information to determine group classification for malingerers and non-malingerers or at the very least provide additional support to increased variability and reaction time rates.

Detection using cut-off scores

Finally, cut-off scores were examined to determine appropriate specificity and sensitivity ratings. When cut-off scores were used an 80% sensitivity rating occurred.

This means that 80% of the participants were correctly identified as malingering on each test. A specificity rating indicates a proper classification of individuals who are not malingering. A specificity of 100% was obtained on each test for the control participants. Although a 93% specificity rating for the Mild TBI Group occurred across all CTIP tasks, the specificity rating for the Moderate TBI Group decreased as the task complexity increased (simple = 93%, choice = 73% and semantic = 47%). This increase in specificity is due to greater information processing impairments for moderate TBI patients. Because the semantic CTIP is a sensitive measure of head injury, specificity for the Moderate TBI Group decreased as complexity increased. Thus, the Moderate TBI Group obtained slower reaction times as complexity increased, and as a result their scores became more similar to those in the Malingering Group. Again, as mentioned previously, the study was less concerned with Moderate TBI ratings as individuals are more likely to malingering a Mild TBI due to insufficient evidence that may occur for this level of injury. It's important to note that a lower sensitivity rating and a higher specificity rating are not always undesirable. In this case, it is less damaging to let someone suspected of malingering go undetected than misclassify someone who has a bona fide injury as 'faking'.

The use of the CTIP as a viable malingering test seems reasonable. It is clear that individuals attempting to malingering produce longer reaction time scores and increased variability than all other groups. As opposed to conventional methods that are becoming more known, and therefore less viable, the CTIP is a new way of detecting malingering. This study could be improved upon in several ways. First, the use of incentives may entice and ensure individuals in the Malingering Group to try harder at going undetected

and therefore may be more representative of a Malingering Group. Second, the administration of the CTIP tasks could be presented in a different order. Alternatively, administering just one of the CTIP tasks amongst other neuropsychological tests may provide different results. Using these alternative test procedures may determine if all CTIP tasks are necessary, if these tests should be given in a certain order, or if just administering a particular task is sufficient. Third, retesting on different days may also be used to examine stability across testing sessions. In other words, a retest may detect whether individuals attempting to malingering can correctly perceive and recall how they performed on the first test.

In the meantime, future studies should take a serious look into obtaining response times from tests currently being administered. That is, if specialized equipment, such as a computer, is being used for administrative purposes, it requires no additional administering time to gather latencies, which in turn could vastly improve predictability of whether an individual is malingering.

Conventional Tests

Overview

The present study compared three tests used to detect malingering: the TOMM-C, VSVT, and 20-Item. All of these conventional malingering tests traditionally have used accuracy in responding as the primary means of detection. However, another possible way to detect malingering is by examining latency scores. The use of reaction time scores is a subtle and non-invasive way to look at possible exaggerated or fabricated symptoms. Examination of response times in the present study revealed a positive

relationship between latency scores and malingering. That is, longer reaction times were associated with an increased likelihood that an individual was malingering.

Test of Memory Malingering (TOMM)

Results of the current study are consistent with findings from the majority of other research on the TOMM: the Malingering Group obtained significantly fewer correct responses than the Control Group (Bolan et al., 2002; Rees et al., 1998; Tan et al., 2002; Tombaugh, 1996). Although the test achieved a 100% specificity rating, surprisingly a lower than expected sensitivity rating occurred (73%). Past studies (Bolan et al., 2002; Rees et al., 1998) have found a 100% classification rating for both groups. A lower detection rate in the present study could have occurred because the TOMM is becoming more familiar. As previously mentioned, an article in the Boston Globe brought to light the ease with which individuals can obtain information on how to perform on such tests by using the Internet (Barry, 2002). In accordance with this statement, Tan et al. (2002) also mention that lower than normal classification ratings may be occurring due to easy access of information in that 36% of their Malingering participants reported they had surfed the Internet in preparation for testing. The ability to easily access information about the TOMM may have decreased the sensitivity. That is, individuals in the malingering group were able to obtain information of how to respond on the TOMM and in turn fewer of these individuals may have gone undetected.

Lower than expected sensitivity also occurred in the Tan et al. (2002) study. They also randomized university students to either a control or malingering group. They found the TOMM provided the lowest classification rating (80%) when compared to two other malingering tests, the WMT (Green et al., 1996) and VSVT. However, a contributing

factor to this result may have been the use of a pencil and paper version of the TOMM. Although Rees et al. (1998) found no significant differences between administering the pencil and paper version and the computerized version, the Tan et al. (2002) study may differ in that all other tests administered were done via a computer. Therefore, the utility of the TOMM may have been compromised (Tan et al. 2002). Additionally, depending on the order of administration, the TOMM may have been perceived as a malingering test. However, they simply listed the tests they administered and did not elaborate on the order in which they were given. If the TOMM had been the last one administered the other tests may have tipped off the test's intentions and thus, a lower than expected classification rating is not surprising. The utility of each malingering test may have also been compromised in that no other neuropsychological tests were administered (with the exception of the Speed of Capacity of Language Processing test). For this reason, the test's intent may have been exposed.

The TOMM obtained only a 65% sensitivity rating in a study by Heinze and Purisch (2001). They used male participants incapable of standing trial on felony charges and who were currently admitted to hospital for further testing. However, a felony charge is likely based on an insanity plea that has nothing to do with memory. As already mentioned, the TOMM is specific to examining individuals attempting to malingering a memory impairment. Thus, using this test to examine other mental deficiencies is not applicable and results based on such are invalid. Additionally, Heinze and Purisch attempted to identify malingering using a multidisciplinary team that based classification on clinical and behavioral observations. No mention was made as to what clinical evidence was provided. Examples of behavioral observation were (1) overly intense

symptoms and (2) bizarre behavior. Recall that in the introduction of the paper that such observations were deemed to be unreliable in detecting malingerers.

A study by Bolan et al. (2002) assigned 32 community volunteers to either a malingering or a control group. All participants were given three malingering tests: the Amsterdam Short Term Memory Test, the Test of Memory Malingering and the Digit Recognition Test. Prior to testing, participants received a scenario but contrary to the current study no time was allowed for 'mental preparation' as testing began immediately thereafter. In the present study, only 3 participants reported they did not prepare for the testing session indicating most of the participants used the time to prepare for the testing. The lack of preparedness in the Bolan et al. (2002) study may not have allowed participants in the malingering group to obtain information on how to perform nor would there have been time to research the topic of malingering. Thus, the logic that the TOMM is becoming a more familiar test does not hold in this case.

Turning to latency scores as a new or improved method of detecting malingering, the current study found the Malingering Group obtained longer latencies than the Control Group. For both groups the latency scores became faster for each subsequent trial with each group producing different scores. These results are similar to what was found in Experiment 5 in the Rees et al. (1998) study. This result is not surprising for trial two, as the subject had already seen all the pictures once before, but is surprising for the retention trial. One would suspect that on the retention trial individuals attempting to malingering would slow down their response due to beliefs that memory impairments and slowed reaction time occur with a head injury. It may be that the retention interval is not long enough and the task is too simple for the expected decrease to occur. In the Tan et al.

(2002) study participants were asked what impairment they attempted to malingering. Most said memory. In the current study, many participants responded the same way when asked, although many inquired during testing if reaction time was a component. Thus, just as accuracy would be expected to decrease during the retention component, it would be reasonable to assume that response times would also increase (i.e., become slower). However, this was not the case. Therefore, it may in fact be that participants were only truly attempting to malingering what they believe is the primary component of head injury (i.e., memory), and were not attempting to adjust their speed.

In another study, Bolan et al. (2002) used 20 undergraduate students who were randomly assigned to a malingering group or a control group. The Bolan et al. (2002) study found the malingering group was significantly slower in responding than the control group on the TOMM. Their study alludes to the fact that reaction time may not be a good discriminator between groups due to a small overlap in the distribution. In their study, there was no mention of a delay period between reading scenario and testing session. Thus, if in fact, no delay period existed, no 'mental preparedness' could have occurred. For this reason, smaller differences in response time could have resulted due to participants not being able to prepare their response. Also, in the Bolan et al. (2002) study, reaction time on the TOMM was defined as time from presentation of two pictures to the point when the spacebar was pressed. That is, after the participant made his/her response in selecting the left or right picture, the space bar was pressed to confirm his/her answer. The added requirement of pressing the spacebar increased the time to response. In the current study only a left or right key response was required in selecting the stimulus. Thus, the current study may be more representative in its response time results

than those of the Bolan et al. (2002) study in that reaction time was defined when pressing only one key.

The current study found no differences in response times between correct and incorrect answers for the Malingering Group. Individuals attempting to malingering must think about their answer, so whether they answer correctly or incorrectly should not have made a difference in their response times. A general response pattern would likely occur, as thought is required for both answers. Put another way, the motor response of pressing the 'correct' or 'incorrect' key does not add time to the response. It is the decision of whether to correctly or incorrectly respond that determines how much time it will take to respond. For this conventional malingering test, participants responded consistently slower on all responses regardless of whether the answer was correct or incorrect. Thus, the reasoning is different from the categorical (same/different) semantic response. One may ask why this conventional test elicited such a response style and the semantic CTIP task did not. It may be that for the semantic task, more information processing is required. For the semantic task an individual needs to cognitively determine if one word belongs to the category of another word, which takes more information processing than remembering and automatically responding to pictures or words.

20-Item Word Test

For the 20-Item Word Test, the current study found the Malingering Group obtained significantly fewer correct responses than did the Control Group. This result is consistent with past studies (Iverson et al., 1994, 1996), but it differs in its ability to classify participants into their respective groups. In the current study, an 80% cut-off was preferable to that of 90% in that it resulted in a 71.4% sensitivity and 78.5% specificity

classification rating. When a cut-off scores of less than 14 (to eliminate false positives) was used a sensitivity rating of 57% and a specificity rating of 100% was found. Several reasons may be responsible for lower classification ratings on the 20-Item Word Test in the current study. First, the test itself has little clinical validation to make it a reliable test in detecting malingering. Secondly, no definitive cut-off scores seemed to be established thereby reducing consistency between studies. Thirdly, it seemed as though the task was challenging for even the Control Group as some participants commented on its difficulty comparatively speaking to the other tasks they were required to complete. In this sense, the test does not meet the criteria of being perceived as hard but in fact is simple. Finally, since the test seemed more challenging to some of the participants they could have taken it as a 'personal challenge' to do well as opposed to following instructions to 'malingering'.

Iverson et al. (1994) asked 60 psychiatric patients and 60 community health volunteers to either try their best or to malingering on the 21-item test. In addition, 30 memory impaired individuals and 30 memory non-impaired individuals were used for comparison purposes. Both of these groups were asked to try their best. Memory impairment was defined as a certain score on the memory scale of the Luria Nebraska Neuropsychological Battery (Golden, Purisch, & Hammeke, 1985) or a score below 85 on the Verbal or General Memory subscales of the Wechsler Memory Scale-Revised (WMS-R: Wechsler, 1987). They found the malingering group performed significantly worse than those who were asked to try their best (for both Community volunteers and Psychiatric patients). Additionally, individuals attempting to malingering obtained significantly fewer correct responses than individuals with a memory impairment. A cut off scores was established to minimize false positives. When a cut-off score of less than

12 was implemented a 70% sensitivity rating and a 99% and 93% specificity rating for the normal control and memory impaired group was found.

Another study by Iverson et al. (1996) study asked a group of students, a group of psychiatric patients and a group of patients with a variation in memory impairments to either try their best or to malingering on the 21-item test. Memory impairment was again defined with the Luria Nebraska Neuropsychological Battery or a score below 85 on the Verbal or General Memory subscales of the Wechsler Memory Scale-Revised (WMS-R). They found those asked to malingering obtained significantly fewer correct responses than those asked to try their best. Their study used a cut-off score of less than 13 to classify participants into their respective groups. Using this cut-off score resulted in an 80% sensitivity rating and a 100% and 95% specificity rating for the normal and memory impaired groups respectively. This cut-off score was established to eliminate false positives. One may question whether this mix of memory-impaired patients can be generalized. Although their study mentioned the type of diagnoses each of these memory-impaired individuals sustained (i.e., seizure disorder, depression, etc), it could be disputed that one type of impairment may contribute to testing difficulties more so than another. In this respect, individuals with more severe impairments would be easily detected, resulting in a higher sensitivity rating than that of the current study. Unfortunately, severity of injury was not elaborated on the Iverson et al. study.

Another distinguishing factor between the current study and some of the studies by Iverson et al. (1994, 1996) is the mode of presentation of the 21-Item Word Test. In both Iverson et al. (1994,1996) studies, the 21-Item Word Test was administered by auditory means. However, in the current study, words were simultaneously presented

orally by the experimenter and visually via a computer. These differences in sensory presentation could have contributed to the differences found in classification ratings. Additionally, the current study did not follow the original procedure of the 21-Item Word Test that required immediate free recall. That is, the current study did not ask participants to recall words. This procedural difference may have contributed to differences in results between the current study and those obtained by Iverson et al. (1994, 1996).

When response time was examined in the present study, significant differences occurred between the Malingering Group and the Control Group. As with the TOMM, no differences occurred for the Malingering Group when examining latency scores for incorrect and correct responses. To date, no other study has examined the potential use of latency scores on the 21-Item Word Test.

Victoria Symptom Validity Test (VSVT)

The VSVT makes certain assumptions regarding how to classify individuals who attempt to malingering. One of these assumptions is based upon correct number of responses. Similar to other malingering tests, Slick, Hoop, Strauss and Thompson (1997) predict that fewer correct responses lead to an increased likelihood than an individual is malingering. The current study found significant group differences on the VSVT, with the Malingering Group obtaining fewer overall correct responses than the Control group, thereby supporting the first assumption.

A second assumption of the VSVT is based upon examination of item difficulty (easy/hard responses). Here, it is assumed that individuals attempting to malingering will obtain fewer correct hard responses than easy responses. The reasoning behind this

assumption is that hard items are perceived as more challenging to remember than easy items (as discussed in the introduction). The current study supports the second assumption. The Control Group had virtually identical number of correct responses for easy and hard items, while the Malingering Group obtained significantly lower scores on the hard items when compared to the easy items. That is, the Malingering Group obtained significantly fewer correct responses than did the Control Group on both easy and hard items; however, the differences between groups was more substantial for hard items. It seems that classifying individuals into their respective groups can be achieved better when examining correct responses for hard items.

A study by Strauss et al. (2000) examined differences in the VSVT across testing sessions. They randomized university students into either a control or malingering group. Participants were tested on three different testing sessions held one week apart. Instructions were given in an envelope but there was no mention if the participants understood the instructions. For the primary session, examination of accuracy found the control group was more accurate than the malingering group, more so for hard items than for easy items. It was found that each group obtained different scores on each level of difficulty (each time interval), this result being more pronounced for the hard items.

Tan et al. (2002) examined the effectiveness of the VSVT at detecting malingering. As in the current study, they found individuals in the malingering group performed similar to the control group on easy items, but this was not the case for hard items. The malingering group obtained significantly fewer correct responses for hard items than the control group. Future study may want to examine the effects of classification when administering only hard items in that this may reduce testing time.

On the other hand, not administering easy items on the VSVT may take away a reference frame as to how much an individual attempting to malingering should delay and/or give correct/incorrect answers.

The final assumption the VSVT makes is that fewer correct responses are made with each increase in time interval. That is, as time between the initial digit presentation and the forced-choice recognition increases from 5 seconds to 10 seconds to 15 seconds, fewer correct responses are made. This assumption is based on the perception that as time increases it is more difficult to recall information. The results from the current study did not find any differences between trials and thus did not provide support that increasing time had any effect on perceived difficulty of the task.

Why did the present study not find that increasing intervals decreased the number of correct responses? This result may have occurred for a few reasons. First, the manual recommends that the test be administered first, primarily for retesting purposes. That suggestion was not followed as interest in other tests took precedence and no retest session took place. Second, fatigue could have played a large part in our results since the test was administered towards the end of the testing session. It could be that participants reached their motivational limit and no longer tried hard. Third, offering an incentive to individuals that participate may have increased their motivation to go undetected. In this study no additional incentive, besides participation credit, was given.

Although the current study is in agreement with the first two assumptions, sensitivity ratings based on established cut-off scores were extremely low in comparison to what other studies have found. Several reasons can be behind the present studies' inability to support others findings. The first, being the most likely contributor, was that

the VSVT was one of the last tests to be administered. Since the entire session was close to two hours, it could have been that participants were tired and in turn were unable to concentrate. The VSVT is unfortunately a long test to administer and as such it may have been difficult for students to concentrate after such a long test session. Additionally, since all other tests had been administered prior to the VSVT, the simplicity of these tasks may have tipped off the tests' primary intent. Individuals in the Malingering Group likely realized the test was too easy and did well. Finally, the present study may have obtained different results than those of past studies on the VSVT simply due to group composition. That is, no definitive explanation can determine why different results occurred except there may have been differences in subjects tested for each study.

As with the TOMM and the 20-Item, reaction time on the VSVT was examined to determine if it were a viable way to detect malingering. For the VSVT, differences between the Malingering Group and the Control Group were found, as were differences between response intervals when latency scores were examined. Similar to the current study, Tan et al. (2003) found the response time for the Malingering Group was significantly longer than the Control Group, more so on hard items than on easy items. The current study found overall latency scores were mainly affected by easy answers signifying that hard items may be difficult to accurately replicate. Put another way, since easy items are just that, easy, it requires a lazy response that is easily influenced. For hard items, some thinking is required and thus just the process of deciding whether to provide a correct or incorrect response may have been considered and not the speed of responding for individuals in the Malingering Group.

Using reaction times to differentiate a head injured group from a malingering group have produced varied results. For example, a study by Rose et al. (1995) administered a computerized PDRT. Individuals attempting to malingering were given an envelope that contained instructions but again there was no mention if the participants understood the instructions. The current study also had participants read a scenario but students were asked to reiterate what they thought they must do to accomplish the written instructions. No confirmation was given but the experimenter was able to determine if appropriate steps and comprehension of instructions were accomplished.

As with the present study, Rose et al. (1995) found significantly longer response times for the malingering group when compared to the control group. However, one unexpected finding was that their head injured group took longer in responding than the malingering group. This result likely occurred due to the severity of the head injury, as individuals who have sustained such injuries generally take longer to respond. In their study, patients had sustained at least a moderate head injury and in turn these patients had sustained more information processing impairment that would have occurred in a Mild TBI group. It is speculated that a different result would likely have occurred if they had used individuals with mild head injuries. So, although one may be able to classify individuals malingering from controls, longer response times could not be a deciding factor between those with a head injury and individuals attempting to malingering in the Rose et al. study.

A study by Strauss et al. (2001) examined response time on the VSVT and found the Malingering group obtained longer response times than the Control group, more so for hard items than for easy items. However, unlike the current study, Strauss et al. found

differences occurred between groups at each response time interval. A highly probable reason for this discrepancy is, yet again, the fatigue factor. The lack of motivation could have affected their performance in that they were unable to slow down in all domains of testing. Meaning, it may have been too much to ask individuals in the Malingering Group to differentiate their response style within trials. Individuals attempting to malingering may have primarily focused on how to respond (correct/incorrect and fast/slow responses) to hard items versus easy items and ignored the increase in time for interval. In addition, as already mentioned, by the time the VSVT was administered a variety of other neuropsychological tests had been employed and this could have ‘tipped off’ the participants. The Malingering Group seemed to differentiate their response style on hard and easy answers for the VSVT. Here, the Malingering Group responded slower to hard items than for easy items. However, when looking across trials, these same subjects generalized their responses. That is, response times for 5-second intervals were similar to response times for both the 10-second and 15-second response intervals. It may be asking too much of subjects to focus on the difficulty (hard/easy) and increased interval times. In this respect, further investigation seems warranted to determine if it is necessary to administer all trials when response times are examined. For example, administering only the 5-second interval on the VSVT can dramatically reduce testing time. This not only decreases the time needed for the participant to be tested but can also reduce costs to the company/practitioner administering the test.

As was not the case with the TOMM-C and the 20-Item, differences between correct and incorrect response time were found. Therefore, it seems that individuals attempting to malingering on the VSVT varied their response style, as was the case with the

semantic test on the CTIP. No explanation can be provided as to why individuals attempting to malingering provided a consistent response style for the TOMM and the 20-Item and not for the VSVT. It may be the more detailed a test is the more likely an individual will differentiate his/her response style.

Strategies

It seemed as though participants in the Malingering group thought they were moderately convincing at representing an individual with brain damage. For the CTIP tests, it seems that this notion holds true, that Malingering individuals perceived they were moderately successful in faking a bona fide brain injury. This judgment perception may strengthen the case that the CTIP tasks may indeed be a new way to examine attempts to malingering.

The issue of students adequately preparing is brought forth with the finding that 3 participants did not prepare at all for the testing session. Again, this may be an indicator that more of an incentive is required to increase motivation to do well. On the other hand, it was found that a large proportion of the students who did prepare for the testing session used the Internet as a means to obtain information. This method of preparedness may support the notion that easy access of information is becoming more apparent. Further research may want to examine what exact information was obtained via the Internet.

Future Research

Exploration into response times seems warranted as the traditional method of detecting malingering may be overused and in turn become more known. Continually using traditional methods in detecting malingering will no doubt become obsolete as

more information becomes accessible to those attempting to malingering a head injury. Already it seems as though these conventional tests are becoming known as shown by their lower sensitivity ratings. The current study shows promise in using response time as a means to detect malingering. Examining each test further showed that for the TOMM-C and the 20-Item, participants do not differentiate their response style within these test. Conversely, for the same/different categorical response within the semantic CTIP task participants were found to differentiate their responses when latency scores were examined.

In summary, future research may want to improve upon the current study. For example, although the present study did not use a TBI group when examining conventional malingering tests, future studies may attempt to use this clinical group for comparison purposes. Several factors could be improved upon. These improvements include the following: an incentive to motivate the participant, randomizing the test administration, reducing the length of the testing session, obtaining a larger sample, and finally using a more representative malingering group than just university students. In the meantime, future studies should take a serious look into latency scores as a new or improved way of detecting malingering.

References

- Alexander, M.P. (1995). Mild traumatic brain injury: Pathophysiology, natural history, and clinical management. Neurology, *45*, 1253-1260.
- American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: American Psychiatric Association .
- Andrikopoulos, J. (1994). Disproportionate attention deficit in malingering.
Poster presentation at the annual meeting of the International Neuropsychological Society, Cincinnati, OH.
- Aubrey, J.B., Dobbs, A.R., & Rule, B.G. (1989). Laypersons' knowledge about the sequelae of minor head injury and whiplash. Journal of Neurology, Neurosurgery, and Psychiatry, *52*, 842-846.
- Barry, E. (2002, June 15). Study finds web of deceit on mental tests. Boston Globe, pp. A1.
- Ben-Porath, Y.S. (1994). The ethical dilemma of coached malingering research. Psychological Assessment, *6*(1), 14-15.
- 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*(5), 715-728.
- Bernard, L.C., Houston, W., & Natoli, L. (1993). Malingering on neuropsychological memory tests: Potential objective indicators. Journal of Clinical Psychology, *49*(1), 45-53.
- Bianchini, K.J., Mathias, C.W., & Greve, K.W. (2001). Symptom Validity Testing: A critical review. Clinical Neuropsychologist, *15*(1), 19-45.

- Bigler, E.D. (1990). Neuropsychology and malingering: Comment on Faust, Hart, and Guilmette (1988). Journal of Consulting and Clinical Psychology, 58(2), 244-247.
- Binder, L.M. (1986). Persisting symptoms after mild head injury: A Review of the postconcussive syndrome. Journal of Clinical and Experimental Neuropsychology, 8(4), 323-346.
- Binder, L.M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. Journal of Clinical and Experimental Neuropsychology, 15(2), 170-182.
- Binder, L.M., & Pankratz, L. (1987). Neuropsychological evidence of a factitious memory complaint. Journal of Clinical and Experimental Neuropsychology, 9(2), 167-171.
- Binder, L.M., & Willis, S.C. (1991). Assessment of motivation after financially compensable minor head trauma. Psychological Assessment, 3(2), 175-181.
- Bink, P.G., Gouvier, D., & Waters, W.F. (1997). Malingering detection with the Dot Counting Test. Archives of Clinical Neuropsychology, 12(1), 41-46.
- Bolan, B., Foster, J.K., Schmand, B., & Bolan, S. (2002). A comparison of three tests to detect feigned amnesia: The effects of feedback and the measurement of response latency. Journal of Clinical and Experimental Neuropsychology, 24(2), 154-167.
- Collins, L.F., & Long, C.J. (1996). Visual reaction time and its relationship to neuropsychological test performance. Archives of Clinical Neuropsychology, 11(7), 613-623.

Davidson, H., Suffield, B., Orenczuk, S., Nantau, K., & Mandel, A. (1991, February).

Screening for malingering using the memory for Fifteen Items Test (MFIT).

Poster presented at the 19th International Neuropsychology Society, San Antonio, TX.

Doss, R.C., Chelune, G.J., & Naugle, R.I. (1999). Victoria Symptom Validity Test:

Compensation-seeking vs. non-compensation-seeking patients in a general clinical setting. Journal of Forensic Neuropsychology, 1(4), 5-20.

Faust, D., & Ackley, M.A. (Ed.). (1998). Detection of malingering during head injury litigation. New York: Plenum Press.

Faust, D., Hart, K., Guilmette, T.J., & Arkes, H.R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. Professional Psychology: Research and Practice, 19(5), 508-515.

Franzen, M.D., Iverson, G.L., & McCracken, L.M. (1990). The detection of malingering in neuropsychological assessment. Neuropsychology Review, 1(3), 247-279.

Fredrick, R.I., & Foster, G.H. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. Psychological Assessment, 3(4), 596-602.

Golden, C.J., Purisch, A.D., & Hammeke, T.A. (1985). Luria-Nebraska Neuropsychological Battery: Forms I and II. Los Angeles: Western Psychological Services.

Gouvier, W.D., Prestholdt, P.H., & Warner, M.S. (1988). A survey of common misconceptions about head injury and recovery. Archives of Clinical Neuropsychology, 3, 331-343.

- Green, P.G., Iverson, G.L., & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. Brain Injury, 13(10), 813-819.
- Greiffenstein, M.F., Baker, W.J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. Psychological Assessment, 6(3), 218-224.
- Guilmette, T.J., Hart, K.J., & Guiliano, A.J., & Leininger, B.E. (1994). Detecting simulated memory impairment: Comparison of the Rey fifteen-item test and the Hiscock forced choice presentation. The Clinical Neuropsychologist, 8(3), 283-294.
- Guilmette, T.J., Sparadeo, F.R., Whelihan, W., & Buongiorno, G. (1994). Validity of neuropsychological test results in disability evaluations. Perceptual and Motor Skills, 78, 1179-1186.
- Hiscock, M., & Hiscock, C.K. (1989). Refining the forced-choice method for the detection of malingering. Journal of Clinical and Experimental Neuropsychology, 11(6), 967-974.
- Iverson, G.L., & Binder, L.M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. Journal of Head Trauma Rehabilitation, 15(2), 829-858.
- Iverson, G.L., Franzen, M.D., & McCracken, L.M. (1991). Evaluation of an objective assessment technique for the detection of malingering memory deficit. Law and Human Behavior, 15, 667-676.
- Larrabee, G.J. (1990). Cautions in the Use of neuropsychological evaluation in legal settings. Neuropsychology, 4, 239-247.
- Larrabee, G.J. (1992). On modifying recognition memory tests for detection of

- malinger. Neuropsychology, 6(1), 23-27.
- Leth-Steensen, C., King Elbaz, Z., & Douglas, V. (2000). Mean Response Time, Variability, and Skew in the Responding of ADHD Children: A response time distribution approach. Acta Psychologica, 104, 167-190.
- Lezak, M.D. (1983). Neuropsychological Assessment (2nd ed.). New York: Oxford University Press.
- Martin, H.J.S., Smirolto, R., Stafford, C., & Gouvier, W.D. (1994). Influence of prior knowledge on the ability to feign mild head injury symptoms in head injured and non head injured college students. Paper presentation at the annual conference of the National Academy of Neuropsychology, Orlando, FL.
- Miller, L. (1996). Malingering in mild head injury and the postconcussion syndrome: Clinical, neuropsychological, and forensic considerations. Journal of Cognitive Rehabilitation, 6-17.
- Millis, S.R., & Volinsky, C.T. (2001). Assessment of response bias in mild head injury: Beyond malingering tests. Journal of Clinical and Experimental Neuropsychology, 23(6), 809-828.
- Nies, K.J., & Sweet, J.J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. Archives of Clinical Neuropsychology, 9(6), 501-552.
- Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. Journal of Consulting and Clinical Psychology, 47(2), 409-410.
- Pankratz, L. (1983). A new technique for the assessment and modification of feigned

- memory deficit. Perceptual and Motor Skills, 57, 367-372.
- Pankratz, L., & Erickson, R.C. (1990). Two Views of Malingering. Clinical Neuropsychologist, 4(4), 379-389.
- Prigatano, G.P., & Amin, K. (1993). Digit Memory Test: Unequivocal cerebral dysfunction and suspected malingering. Journal of Clinical and Experimental Neuropsychology, 15(4), 537-546.
- Rees, L.M., Tombaugh, T.N., Gansler, D.A., & Moczynski, N.P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). Psychological Assessment, 10(1), 10-20.
- Resnick, P.J. (1988). Malingering of posttraumatic disorders. In R. Rogers (Ed.), Clinical assessment of malingering and deception (pp. 84-103). New York: Guilford Press.
- Rogers, R. (Ed.). (1988). Clinical assessment of malingering and deception. New York: Guilford Press.
- Rose, F.E., Hall, S., & Szalda-Petree, A.D. (1995). Portland Digit Recognition Test – Computerized: Measuring response latency improves the detection of malingering. Clinical Neuropsychologist, 9(2), 124-134.
- Ruff, R.M., Wylie, T., & Tennant, W. (1993). Malingering and malingering-like aspects of mild closed head injury. Journal of Head Trauma Rehabilitation, 8(3), 60-73.
- Schacter, D.L. (1986). On the relation between genuine and simulated amnesia. Behavioral Sciences & the Law, 4(1), 47-64.
- Schmand, B., de Sterke, S., Lindeboom, J., & Millis, S.R. Amsterdam Short Term

Memory Test Manual (in preparation).

Schmidt, J.P., & Tombaugh, T.N. (1995). Learning and Memory Battery (LAMB).

Toronto, Canada: Multi-Health Systems Inc.

Schretlen, D., Brandt, J., Krafft, L., & Van Gorp, W. (1991). Some caveats in using the re 15-item memory test to detect malingering amnesia. Journal of Consulting and Clinical Psychology, 3(4), 667-672.

Slick, D., Hoop, B., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's Forced-Choice Memory Test. Journal of Clinical and Experimental Neuropsychology, 16(3), 472-481.

Slick, D., Hoop, B., Strauss, E., & Thompson, G.B. (1997). The Victoria Symptom Validity Test. Odessa: PAR.

Slick, D.J., Sherman, E.M.S., & Iverson, G.L. (1999). Diagnostic criteria for malingering neurocognitive dysfunction: Proposed standards for clinical practice and research. Clinical Neuropsychologist, 13(4), 545-561.

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. Clinical Neuropsychologist, 14(4), 420-432.

Strauss, E., Slick, D.J., Levy-Bencheton, J., Hunter, M., MacDonald, S.W.S., & Hultsch, D.F. (2001). Intraindividual variability as an indicator of malingering in head injury. Archives of Clinical Neuropsychology, 17, 423-444.

Strauss, E., Spellacy, F., Hunter, M., & Berry, T. (1994). Assessing believable deficits

- on measures of attention and information processing capacity. Archives of Clinical Neuropsychology, 9(6), 483-490.
- Tan, J.E., Slick, D.J., Strauss, E., & Hultsch, D.F. (2002). How'd They Do It? Malingering Strategies on Symptom Validity Tests. The Clinical Neuropsychologist, 16(4), 495-505.
- Tombaugh, T.N. (1996). The Test of Memory Malingering (TOMM). Toronto, Canada: Multi-Health Systems Inc.
- Tombaugh, T.N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individual. Psychological Assessment, 9(3), 260-268.
- Tombaugh, T.N. (1999). The Adjusting-Paced Serial Addition Test (Adjusting-PSAT). Unpublished research, Carleton University.
- Tombaugh, T.N. (2001). The Test of Memory Malingering (TOMM) in Forensic Psychology. Manuscript submitted for publication.
- Tombaugh, T.N. and Rees, L. (1999). Computerized Tests of Information Processing. Unpublished test. Carleton University, Ottawa Ontario.
- Trueblood, W., & Binder, L.M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. Archives of Clinical Neuropsychology, 12(1), 13-27.
- 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(4), 578-590.
- Vallabhajosula, B., & van Gorp, W.G. (2001). Post-Daubert admissibility of scientific

- evidence on malingering of cognitive deficits. Journal of American Academy of Psychiatry and the Law, 29(2), 201-215.
- Van Gorp, W.G., Humphrey, L.A., Kalechstein, A., Brumm, V.L., McMullen, W.J., Stoddard, M., & Pachana, N.A. (1999). How well do standard clinical neuropsychological tests identify malingering? A preliminary analysis. Journal of Clinical and Experimental Neuropsychology, 21(2), 245-250.
- Van Zomeren, A.H., & Brouwer, W.H. (1994). The assessment of attention. In clinical neuropsychology of attention (pp. 159-178). New York: Oxford University Press.
- Vickery, C.D., Berry, D.T.R., Inman, T.H., Harris, M.J., & Orey, S.A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. Archives of Clinical Neuropsychology, 16(1), 45-73.
- Wechsler, D.A. (1987). Wechsler Memory Scale-Revised manual. New York: Psychological Corporation.
- Wiggins, E.C., & Brandt, J. (1988). The detection of simulated amnesia. Law and Human Behavior, 12(1), 57-78.
- Wogar, M.A., Broek, M. D., Bradshaw, C. M., & Szabadi, E. (1998). A new performance-curve method for the detection of simulated cognitive impairment. British Journal of Clinical Psychology, 37(3), 327-341.
- Zielinski, J.J. (1994). Malingering and defensiveness in the neuropsychological assessment of mild traumatic brain injury. Clinical Psychology: Science and Practice, 1(2), 169-183.

Appendix A

Questionnaire

1. Please indicate how hard you tried to follow the instructions given at the beginning of the experiment by circling a number that described your best efforts.

1	2	3	4	5
didn't try at all		tried moderately hard		tried very hard

2. What resources did you use in preparing for today's testing?

☐ Books, TV, course notes
☐ Internet
☐ Asked friends, family
☐ Asked Professional
☐ Nothing
☐ Other: _____

3. Please indicate how successful you feel you were at following the instructions by circling a number.

1	2	3	4	5
not very		moderately		very

4. Do you have any personal knowledge and/or experience with brain injury?

☐ No
☐ Yes: family, friends, professional (i.e. doctor, professor, etc.)
☐ Yes: myself
☐ Yes: courses, TV, books
☐ Yes: Internet
☐ Yes: other (please list): _____

How convincing do you think you were in representing an individual with brain damage on each of the tests administered? Please place a check on the scales below that best represents how well you think you did.

	Not Convincing 0%	50%	Very Convincing 100%
1. Pressing a key when an X appeared on the screen.	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
2. Pressing a key for words "duck" and "kite".	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
3. Pressing a key for semantic associations .	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
4. Recognizing pictures of 50 objects.	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
5. Selecting 5 digit numbers	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
6. Recalling/recognizing 20 noun word list.	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
7. Repeating a series of numbers forward and backwards.	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
8. Connecting circles containing numbers and letters.	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
9. Matching numbers to symbols	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		
10. Adding consecutive digits together.	<div style="border-top: 1px solid black; position: relative; height: 15px;"> </div>		

Please rate the tests below in there ability to catch someone trying to fake a brain injury
(1 = best to 10 = worst).

1. Pressing a key when an
X appeared on the screen. _____
2. Pressing a key for words
“duck” and “kite”. _____
3. Pressing a key for semantic
Associations. _____
4. Recognizing pictures of
50 objects. _____
5. Selecting 5 digit numbers _____
6. Recalling/recognizing
20 noun word list. _____
7. Repeating a series of numbers
forward and backwards. _____
8. Connecting circles containing
numbers and letters. _____
9. Matching numbers to
Symbols. _____
10. Adding consecutive digits
together. _____

1. What, if any, steps did you take in preparation for taking the test?

2. Did you use a specific strategy for each test: yes or no (circle).

(a) If yes, what strategy did you use?

(b) If no, please indicate what different strategies you used on each of the test:

1. Pressing a key when an X appeared on the screen
2. Pressing a key for words “duck” and “kite”
3. Pressing a key for semantic associations
4. Recognizing pictures of 50 objects
5. Selecting 5 digit numbers
6. Recalling/recognizing 20 noun word list
7. Repeating a series of numbers forward and backwards
8. Connecting circles containing numbers and letters
9. Matching numbers to symbols
10. Adding consecutive digits together

3. Circle the symptom or symptoms you feel that one is most likely to exhibit when attempting to fake a brain injury:

Memory Impairment

Attention