Validity of VSVT and Reliable Digit Span Measures for Detecting Malingering in Groups Differing in Exposure to and Knowledge of TBI

by

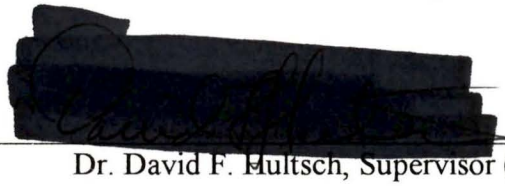
Judith Levy-Bencheton
B.A. Honours, York University, 1998

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

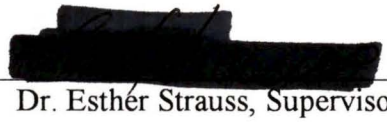
MASTER OF SCIENCE

in the Department of Psychology

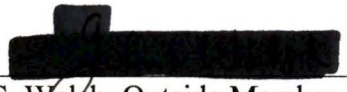
We accept this thesis as conforming
to the required standard



Dr. David F. Hultsch, Supervisor (Department of Psychology)



Dr. Esther Strauss, Supervisor (Department of Psychology)



Dr. John C. Walsh, Outside Member (Department of Psychological Foundations, Faculty of Education)

Dr. Zheng Wu, External Examiner, (Department of Sociology)

University of Victoria

Supervisors: Drs. David F. Hultsch and Esther Strauss

ABSTRACT

Recent empirical findings (Strauss et al., 1999) demonstrated the potential utility in examining variability in performance to diagnose malingering using a standard neuropsychological measure and a forced choice symptom validity task. However, Strauss et al. (1999) used an analog paradigm in which they did not examine the potential contribution of knowledge of head injuries (through direct experience with head-injured individuals or having received a head injury oneself) to malingering status. The main purposes of this thesis include providing cross-validation for the neuropsychological measure and the Victoria Symptom Validity Task (VSVT) in detecting malingering status, examining the contribution of variability in performance to detecting malingering, as well as examining of the impact of knowledge on malingering status. A sample of 74 participants (27 head-injured adults, 26 naïve people, and 21 experienced individuals) was recruited from advertisements and newspaper ads. The participants ranged in age from 18 to 57 years ($\underline{M} = 36.24$, $\underline{SD} = 9.68$). All participants provided demographic and health-related information. They also completed a battery of cognitive tests including the MMSE (Folstein, Folstein, & McHugh, 1975), the Wechsler Adult Intelligence Scale-III (WAIS-III) Block Deign and Vocabulary subtests (Psychological Corporation, 1997), and the North American Adult Reading Test (NAART, Blair and Spreen, 1989) in order to obtain benchmark cognitive measures. The study took place across four occasions, consisting of an initial interview and three test sessions. Following the initial interview, participants were randomly assigned to one of two experimental groups: malingerers and controls. Participants were tested individually over three measurement occasions

scheduled approximately two weeks apart. The dependent measures of interest included the VSVT and the RDS.

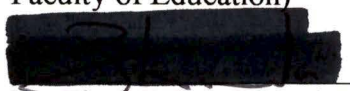
The results indicated that multiple measures of malingering (VSVT and RDS) obtained in a single assessment (occasion one, two or three) discriminated the groups effectively. In addition, however, intraindividual variability in performance (on both tasks) provided unique information beyond level of performance. There was no effect of knowledge in detecting malingering status. The results suggest that response inconsistency across testing sessions may be a clinically useful measure for the detection of malingering.

Examiners:

Dr. David F. Hultsch, Supervisor (Department of Psychology)

Dr. Esther Strauss, Supervisor (Department of Psychology)

Dr. John C. Walsh, Outside Member, (Department of Psychological Foundations, Faculty of Education)



Dr. Zheng Wu, External Examiner, (Department of Sociology)

Table of Contents

Page

Table of Contents (cont.)

Page

List of Tables

Page

List of Figures

Page

Acknowledgements

To begin with, I would like to thank my supervisors Dr. David F. Hultsch and Dr. Esther Strauss for their support and guidance throughout the entirety of this project. Their professional guidance, and personal support were an asset in running the study and in writing this thesis. It has also provided me with a foundation on which to build from in my future endeavours. In addition, I would like to thank Dr. John C. Walsh for his excellent suggestions and his understanding throughout this process.

In addition, I would like to express my sincere gratitude to some senior graduate students who have supported me both professionally and personally. First, I would like to thank Stuart MacDonald for his guidance throughout this whole process, both in executing the study and in analyzing the results. Your patience, and ability to mentor both through the academic trials and personal frustrations associated with running a study and writing a thesis are deeply appreciated. To Sue Larke, your personal support and willingness to diligently peruse my draft offering excellent suggestions have been integral to my finishing this thesis.

On a personal level, I would like to thank my family and friends both here and in Toronto for their understanding, encouragement and belief in my abilities. My parents, Sheila and Prosper Levy-Bencheton deserve special thanks for listening to me on almost a daily basis and unconditionally providing an understanding and caring ear for me.

Acknowledgements

To begin with, I would like to thank my supervisors Dr. David F. Hultsch and Dr. Esther Strauss for their support and guidance throughout the entirety of this project. Their professional guidance, and personal support were an asset in running the study and in writing this thesis. It has also provided me with a foundation on which to build from in my future endeavours. In addition, I would like to thank Dr. John C. Walsh for his excellent suggestions and his understanding throughout this process.

In addition, I would like to express my sincere gratitude to some senior graduate students who have supported me both professionally and personally. First, I would like to thank Stuart MacDonald for his guidance throughout this whole process, both in executing the study and in analyzing the results. Your patience, and ability to mentor both through the academic trials and personal frustrations associated with running a study and writing a thesis are deeply appreciated. To Sue Larke, your personal support and willingness to diligently peruse my draft offering excellent suggestions have been integral to my finishing this thesis.

On a personal level, I would like to thank my family and friends both here and in Toronto for their understanding, encouragement and belief in my abilities. My parents, Sheila and Prosper Levy-Bencheton deserve special thanks for listening to me on almost a daily basis and unconditionally providing an understanding and caring ear for me.

Dedication

To my parents, whose love and support guides me through all of my endeavours.

Chapter I

INTRODUCTION

In recent years both the health care and legal systems have been increasingly concerned with assessing the veracity of individuals' claims of poor health in psychological functioning. In neuropsychology, particular concern has been directed toward the development of instruments designed to detect malingering following putative traumatic brain injury. Although the precise prevalence of malingering after head injury is unknown, preliminary estimates range from 15% to 64% depending on the population studied (Hayes, Hilsabeck, & Gouvier, 1999) (e.g., workers' compensation applicants, personal injury litigants, and Social Security Administration disability claimants). The significant incidence of malingering, coupled with an estimated 2 million individuals seeking medical attention annually as a result of a closed head injury, necessitates an understanding and ability to diagnose malingering from true behavioural problems resulting from head injury.

The American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (1994) defines malingering as "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation...." This definition of malingering encompasses both the exaggeration of preexisting symptoms and the feigning of symptoms that are nonexistent. Because head injuries are often compensated if they happen at work or in vehicular accidents, the falsification of related symptoms is a substantial concern that can result in an unjust distribution of monetary compensation to bona fide patients and malingerers.

Neuropsychologists are often hesitant to diagnose malingering due to the ethical and legal implications associated with the designation and the implicit understanding that the relationship between psychologist and patient should be both honest and cooperative (Binder, 1990). Malingering does become a diagnostic consideration whenever (a) readily identifiable and commonly recognized incentives for exaggeration exist, (b) subjective complaints or test results are not consistent with neurological or functional status, (c) symptoms and complaints do not make medical sense, (d) there is a history of emotional/ personality disorder (e.g., sociopathic behavior), and/ or (e) patient cooperation is questionable (Strauss et al., 1999). These diagnostic criteria for malingering incorporate a variety of variables that aid in accurately identifying those who are feigning cognitive dysfunction from those who are not. This hesitancy in diagnosing malingering emphasizes the need for valid and reliable assessment measures to accompany the aforementioned factors in order to provide the most accurate diagnosis possible.

Tasks designed to detect malingering have an additional factor to account for that one typically does not consider in neuropsychological tests. The standard assumption that the client is performing at their optimal level (from the examiner's point of view) is not applicable in situations where malingering is a possibility. In addition to disregarding this assumption, one has to consider that knowledge of the current means to assess malingering could potentially impact optimal performance on neuropsychological tests. Due to the adversarial and competitive nature of litigation, legal professionals have become increasingly sophisticated in their knowledge of the current measures used to detect malingering. In addition, a recent survey (Wetter & Corrigan, 1995) reported that

80 percent of practicing attorneys (n=80) believed that they should inform their clients about psychological testing, and half of these individuals believed that this included educating the plaintiff about symptom validity scales. Youngjohn, Lees-Haley, and Binder (1999) confirmed the preponderance of coaching that occurs in medicolegal contexts by reviewing the literature on coaching, and the positive effect that it has on individuals eluding detection. Consequently, health professionals must find ways to circumvent knowledge that a patient may have, and provide an accurate diagnosis based on effective tasks that distinguish between malingering and genuine cognitive deficits.

Malingering and Mild Traumatic Brain Injury (TBI)

The diagnostic relevance of malingering for those who have suffered a mild TBI is of considerable concern due to the high incident rate of head injuries (Binder, 1997; Hayes et al., 1999), and the ambiguity associated with their associated symptomology. Post-concussive syndrome (PCS) is often used to describe the symptoms experienced by individuals after a head injury. PCS covers a wide array of symptoms including physical problems (e.g., dizziness, fatigue, nausea, and headaches), cognitive complaints (e.g., difficulty in concentrating, problems with memory), and behavioural and affective symptoms (e.g., irritability, anger, depressed mood) (Bernstein, 1999). The nonspecificity of these symptoms makes it difficult to distinguish PCS from other problems with different etiologies such as anxiety and depression. There are studies revealing this cluster of symptoms as being no less common among controls than in head-injury individuals (Gouvier, Hayes, Smiroldo, 1998), thus demonstrating the commonality of these symptoms and the possibility of a wide array of etiologies.

There seems to be little empirical evidence to indicate that prolonged neuropsychological deficits are caused by mild TBIs. A recent review by Binder (1997) suggests that most individuals who have suffered a mild head injury recover approximately 3-6 months post injury. A small percentage of patients persist with complaints and it is this group that clinicians have to deal with and make a determination concerning the individual's diagnosis. Binder has suggested that when symptoms persist for more than 6 months, alternative explanations must be considered independent of the head injury. Indeed, there has been controversy in the literature as to whether or not PCS can persist past 6 months (see Bernstein, 1999).

Due to this ambiguity associated with the outcome of a mild head injury, it becomes extremely difficult to dissociate individuals complaining of genuine complaints from those that are feigning injury. The vague symptomology associated with PCS does not provide adequate criteria for diagnosing individuals after mild head injury. This renders it extremely important to improve diagnostic methods of malingering after head injury. The current research uses different paradigms to investigate the utility of several malingering measures.

Analog Versus 'Real Life' Studies

Malingering is a difficult issue to address empirically. There is a myriad of reasons that individuals may be malingering on both a conscious and unconscious level. This makes it difficult for the clinician to predict malingering behaviour. Additionally, one should consider the fact that if the individual is malingering, they will most often not state this explicitly. Malingering research focuses on how to distinguish malingerers from non-malingerers without the confirmation of explicit admission.

Iverson and Franzen (1996) identify two general research methodologies that are used to study malingering: i.e. analog studies conducted in the laboratory, and 'real life' studies using litigating and non-litigating head injury patients. In the analog approach, participants are asked to simulate or feign a head injury for monetary compensation. This is likened to the individual who malingers while seeking compensation from the insurance companies. The potential flaw with this design is a lack of ecological validity of these studies. To what degree are the findings generalizable to the public given that the simulators did not actually experience a head injury and typically will not obtain any significant monetary compensation for malingering successfully?

The other methodology in malingering research uses a real life context. One way of accomplishing this is to employ probable malingerers (from hospitals and private practices) as participants. These patients are identified as such by criteria set out in the study (e.g., Martin, Franzen, Orey, 1998; Van Gorp et al., 1999). Practical issues associated with this research arise such as the ability to identify individuals as true malingerers, as well as small sample sizes due to the difficulty in attaining clinical populations in research studies. Another method of creating a real life context is by contrasting head-injured litigants to those not in litigation (e.g., Meyers & Volbrecht, 1998). This method incorporates the impact of the legal system on malingering tests. Coaching effects, and incentives to malinger are theoretically present for the litigants and absent for the nonlitigants. The apparent flaw with this research is that one can not ascertain who is malingering versus who is not on the sole basis of being in litigation. Because of the associated difficulties with both methods of using a real life context, most research is done in a laboratory setting. Efforts are made to increase ecological validity

by coaching subjects and offering incentives for successful malingering and where possible, using individuals who have suffered head injuries, and have medical documentation of the severity of the injury.

There have been studies attempting to increase the validity of analog designs by trying to recreate real life settings. Rose, Hall, Szalda-Petree, and Bach (1998) examined TBI and malingering in an analog study. One of their groups of malingerers was simply told to fake deficits to the best of their ability while the other group was given information on head injury, much the same as is in litigation where a lawyer or client may investigate the effects of head injury on cognitive performance. Their control groups consisted of naïve, healthy individuals as well as a second group of head-injured individuals. Rose et al. (1998) found that participants who were coached were able to avoid malingering diagnoses better than their uncoached counterparts. Other studies have yielded similar results (e.g., Martin, Gouvier, Todd, and Bolter, & Niccolls, 1992). Such findings emphasize the need to develop paradigms that are robust to coaching effects.

Other efforts to improve analog studies include offering financial incentives for 'successful' feigning (e.g., Demakis, 1999), as well as using participants that have suffered genuine head injuries or have complained of memory problems, independent of any legal context (e.g., Iverson & Franzen, 1996; Slick, Hopp, Strauss, Spellacy, 1996). All of these efforts could improve classification of malingerers, but there are currently no available studies that focus on determining the most optimal context in which to evaluate malingering. This would necessitate the inclusion of a multitude of factors including false positives in 'real life' studies, as well as analyzing all potential efforts that constitute increasing the validity of analog studies. Certain efforts have been evaluated in isolation

such as the impact of coaching to increase the validity of the study (Rose et al., 1998), but the combination of evaluating analog versus real life studies, coaching effects, and the impact of the litigation process have not been looked at simultaneously.

The existence of analog and real life studies in the current literature is helpful in examining malingering tasks across different contexts thereby optimizing the validity and reliability of the given tests. Studies comparing and contrasting the approaches to studying malingering would be useful as well and examining the relevant factors impacting malingering performance.

Malingering Tests

In both analog and 'real life' settings, there are two measurement approaches that have been used by psychologists to diagnose malingering. The first method uses standard neuropsychological assessments such as the Wechsler Memory Scale-Revised (Franzen & Martin, 1996), or the Rey Auditory Verbal Learning Test (Bernard, 1991). The second method uses symptom validity testing, a forced choice paradigm in which participants are forced to select a response. On traditional neuropsychological tests, cutoff scores are used in order to assess the veracity of the patient's performance. Because these tests are typically included in any neuropsychological assessment, they are also routinely used to diagnose malingering behaviour. Slick et al., (1996) discuss the lack of cross-validation of these studies while still acknowledging their utility due more to convenience as opposed to the validity of the measures.

Digit Span and traditional neuropsychological tests. One of the traditional neuropsychological tests commonly used as a measure to detect malingering is the Digit Span subtest of the WAIS-III (e.g., Franzen & Martin, 1996; Heaton, Smith, Lehman, &

Vogt, 1978; Iverson & Franzen, 1996; Meyers and Volbrecht, 1998; Mittenberg, Therouz-Fichera, Zielinski & Heilbronner, 1995). A comprehensive neuropsychological examination typically includes the administration of the Wechsler intelligence scales because of the potential effect the head injury has on the patient's intellectual abilities. The utility in using Digit Span specifically as a malingering measure is a function of it being a component in over 90% of all neuropsychological examinations (Mittenberg et al., 1995), thus being readily available to analyze for medical-legal purposes. It can also be quickly administered and has been found to be a useful diagnostic tool for malingering.

Numerous studies have established reductions in digit span performance as characteristic of malingered head injury (Binder & Willis, 1991; Mittenberg et al., 1993). The rationale for this lies in the fact that immediate attention span, measured by Digit Span remains preserved relative to overall cognitive level in concussed patients (Levin et al., 1982; Mittenberg et al., 1993). Individuals feigning injury often suppress their digit span performance, assuming that immediate attention span is a measure of memory thereby not remaining intact after injury. Due to its utility and convenience, the Digit Span task has been used in both analog and real life studies as a malingering measure.

Mittenberg et al. (1995), using an analog paradigm, contrasted a sample of 67 volunteers asked to simulate a head injury with 67 patients who had genuinely sustained a head trauma. All participants were tested with the standardized form of the WAIS-R. The results indicated that head trauma patients showed similar levels of performance on the Vocabulary and Digit Span subtests, but participants feigning injury showed exaggerated reductions on Digit Span compared to their vocabulary scores. Using this

difference score in a discriminant function analysis, a 63% sensitivity rate and a 79% specificity rate were obtained. The authors suggest that in order to increase sensitivity and specificity rates, a patient's neurological status should be determined from a variety of sources.

Analog studies that followed using Digit Span (as well as other neuropsychological measures) as a malingering measure included that of Iverson and Franzen (1996), Franzen and Martin (1996), and Strauss et al. (1999). All studies further substantiated the use of Digit Span as an indicator of malingering while cautioning the use of it as the sole basis for a diagnosis.

The digit span task has also been used in real life studies (e.g., Greiffenstein, Gola & Baker, 1995; Greiffenstein, Gola & Baker, 1994; Suhr, Tranel, Wefel & Barrash, 1997; Trueblood & Schmidt, 1993). These studies, similar to analog paradigms, demonstrated suppressed performance of digit span by probable malingerers. Although the utility of the task has been illustrated, efforts have been made to further strengthen it by extracting different variables from the task, and comparing Digit Span to other cognitive measures.

One problem with measuring malingering with intelligence measures like Vocabulary, or Digit Symbol is that the differences between the concussed and the malingering groups could be secondary to more impaired overall levels of IQ in the concussed group. Mittenberg et al. (1995) sought to rectify this by measuring a reduction in Digit Span relative to Vocabulary, which is most closely related to overall level of intellectual function and typically not depressed because malingerers do not see it as a memory measure (Wechsler, 1981). Thus, if individuals scored high on vocabulary (demonstrating a high level of intellectual function) and considerably lower on digit span,

this was thought to be an indication of suppressed performance. Strauss et al. (1999) further substantiated the utility of the Vocabulary minus Digit Span score by differentiating malingerers from nonmalingerers amongst their sample of college students.

In addition, the Digit Span task can yield different variables that are used in analysis. Greiffenstein et al. (1994, 1995) used a new measure, namely the Reliable Digit Span (RDS). It is calculated by summing the longest string of digits repeated without error over 2 trials under both forward and backward conditions. RDS has since been used in other malingering studies like that of Meyers and Volbrecht (1998).

Meyers and Volbrecht used Reliable Digits (RD) and a forced choice measure to diagnose malingering. One of the main purposes of this study was to cross validate Greiffenstein et al. (1994; 1995). Forty-seven mild brain-injured litigating participants and 47 mild brain-injured nonlitigating participants completed the digit span task as well as a forced choice task. RD classified 4.1% of the non-litigating participants, and 48.9% of the litigating participants as malingerers. RD classified more litigating participants than the forced choice task, and none of the nonlitigating participants failed the forced choice measure. RD was more effective in increasing the sensitivity while the forced choice task increased the specificity rates. This further substantiates the need for more than one measure of malingering in order to make accurate diagnoses.

Given the frequency with which the Digit Span subtest is typically administered, and the information it provides as to whether the individual is malingering or not, it seems like a useful addition to the diagnostic enterprise. It is also important to assess

whether other standard neuropyshological tests are useful in diagnosing malingering, given their frequency in administration.

Van Gorp et al., (1999) have recently addressed the need to examine how well standard neuropsychological tests (e.g., Trail Making Test, Reitan, 1958; Boston Naming Test, Kaplan, Goodglass, & Weintraub, 1983; Stroop Color Interference test, Spreen & Strauss, 1991) differentiated malingerers from non-malingerers. A small sample of mild to moderate head injured individuals was identified on the basis of various diagnostic criteria as probable malingerers, with a larger control group as the non-malingerers. Van Gorp et al. (1999) found that the pattern of neuropsychological tests alone did not suffice in identifying the malingerers. Discriminant function analysis on the pattern of neuropsychological test scores correctly classified 84.0% of the subjects, with 72.2% of the suspected malingerers, and 87.7% of the non-malingerers correctly classified. Cross validation of this analysis revealed that use of data from the clinical neuropsychological tests was no more effective than chance in determining probable malingering versus non-malingering status. Level of performance on each test was a better indicator of malingering, supporting the notion that malingerers 'aim too low' in portraying themselves as impaired. Although this is only a preliminary study based on a small sample, it provides evidence for the need to include measures that are specifically designed to detect malingering. In addition, the study suggests that these measures must be based on the notion that malingerers typically perform at a very low level.

Symptom validity tests. Symptom Validity Testing (SVT) is a method that is specifically focused on detecting malingering behaviour. It is commonly used to assess suboptimal performance in those who have suffered a head injury. SVT refers to a forced

choice paradigm in which participants are asked to select among a finite (e.g. one of two) number of responses. In the presence of cognitive deficits, participants should not score below chance levels (given a reasonable confidence interval) on the test. Anything below chance is suspect as it is possible that the individual knew the answer and purposely chose the wrong answer to feign cognitive deficit. Although this is a viable method to assess malingering, studies have shown that probable malingerers do not always perform at below chance levels (e.g., Rose et al., 1998), rendering symptom validity testing susceptible to error in diagnosing. The use of normative data can be helpful here.

Slick et al. (1996) identify the need to make symptom validity tasks appear as difficult as possible to the malingerer while maintaining the task's feasibility for the patient. This can be done by employing seemingly more complex stimuli, increasing difficult distracter tasks, increasing time intervals during memory tasks, and using deceptive instructions that stress the 'difficulty' of the tasks. The Victoria Symptom Validity Test (VSVT) that is used in this study exemplifies the aforementioned modifications. It is a computerized forced choice task whereby the individual is shown a set of numbers, and later asked to identify the set of numbers from a pair consisting of the target and a foil. The additional possibility of examining reaction time is useful as malingerers often increase their response latencies (Strauss, Spellacy, Hunter & Berry, 1994).

Although symptom validity testing has demonstrated some convincing evidence in predicting group membership (Bickart, Meyer, Connel, 1991; Binder, 1990; Binder et al., 1993; Slick, Hopp, Strauss, Hunter, & Pinch, 1994), there are potential shortcomings to this approach. For example, while attempts are being made to increase the face

validity of the measures by making the tests appear more difficult, the potential of floor effects could decrease the specificity of any normative based cut-off scores developed for these tests (Binder et al., 1993; Iverson, Franzen & McCracken, 1991). That is to say that by decreasing the specificity, one may have difficulty identifying those who are malingering (Gouvier et al., 1998).

Due to the fallibility of individual malingering measures (both in traditional neuropsychological tests and symptom validity tests), the use of multiple measures assessing malingering seems like a plausible method to increase accurate diagnoses (Mittenberg et al., 1995). Iverson and Franzen (1996) documented the performance of experimental malingerers, controls and memory impaired individuals on several measures including symptom validity testing and traditional neuropsychological reports. The symptom validity tests included the 21 Item Test (Iverson et al., 1991), and an addition to the Logical Memory test (a paragraph recall test). Both tests were comprised of a forced choice recognition task that would render any score that was significantly below chance levels as suspect. The traditional neuropsychological tests included the Memorization of 16 Items test (that was developed by Rey (1941) and later modified by Paul, Franzen, Cohen, & Fremouw (1992)), the Digit Span (Weschler, 1981), and the Logical Memory subtest (Weschler, 1987). The Memorization of 16 Items test is a task that requires the memorization of 16 symbols. In reality, patients need only remember three or four of the ideas to recall most of the items. The test actually proves to be one on which the most severely brain damaged or cognitively impaired patients perform easily. The Logical Memory subtest requires the immediate memorization of facts from a story, as well as a delayed recall of the facts. These tests were chosen because they are common tests that

head-injured individuals can typically perform at normal levels, whereas individuals who are malingering tend to exaggerate their deficits. Iverson and Franzen (1996) found that combining their measures improved classification rates over using a single measure, yielding a 92.5% correct classification rate for the experimental malingerers, and a 100% hit rate for the control and memory impaired subjects who were instructed to try their best. This study, along with others (e.g., Meyers and Volbrecht, 1998) substantiates the notion that multiple measures are useful in detecting malingering behaviour.

In addition to the use of multiple measures, maximizing the numbers of dependent variables obtained from one test could aid in the detection of malingering behaviour. From a statistical vantage-point, using more variables in a classification analysis such as logistic regression or discriminant function could potentially improve the model, yielding more accurate classification rates. It has also been suggested that the more variables that are measured in assessments, the less susceptible neuropsychological assessments become to errors in diagnosis (Heubrock, & Peterman, 1998). Martin, Franzen, & Orey (1998) attempted to extract another dependant variable from the available data by examining participants' pattern of errors on the Logical Memory and Visual Reproduction subtests from the Wechsler Memory Scale-Revised (WMS-R; Weschler, 1987) in order to detect malingering behaviour. They found that the malingerers were more likely to select the low probability items on multiple choice recognition tests than controls or head injury patients. Although Martin et al. (1998) used a small sample size, and did not assess compliance and motivation to feign their performance, the study is still important because it used all of the data generated from existing tests to shed further insight upon malingering behaviour. This method could be employed with other traditional tests like

the Wisconsin Card Sorting Test, and the Visual Form Discrimination Test as discussed by Rogers, Harrell, & Liff (1993). It could prove a useful technique that would not require any more time consuming or costly tests for the participant in order to obtain additional information to aid in detecting malingering behaviour. The principle of extracting all of the information possible from one test (given repeatedly) could be implemented beyond the scope of the aforementioned article. In addition to looking at level of performance, variability in performance warrants some investigation to determine if adds unique information beyond level of performance on a given task.

Variability in Performance and Predicting Malingering

Variable performance across a number of sessions could serve as an additional factor in predicting group membership between controls and malingerers. The nature of the legal system frequently necessitates multiple testing sessions of patients suffering from mild TBI (Reitan & Wolfson, 1996). Accordingly, it would be quite feasible to measure variability in performance on neuropsychological tests if it proved to be a useful tool to diagnose malingering.

The notion of variability or change is a complex one to define. Nesselroade (1991) divides within person change into two distinct concepts. Intraindividual change refers to change that is more or less durable and systematic (e.g., learning, development, changes in traits), whereas intraindividual variability refers to change that is more or less transient (e.g. cyclical fluctuations, changes in states, 'wobble'). The literature focusing on variability and malingering operationalizes variability as what Nesselroade refers to as intraindividual change. More specifically, researchers propose that the examination of

variability may provide important diagnostic information since the malingerer may have difficulty sustaining the same level of inadequate effort over multiple test sessions.

Studies that have investigated variability in performance of people who have suffered head injuries (Cullum, Heaton, & Grant, 1991; Demakis, 1999; Reitan &Wolfson, 1995, 1996, 1997;) have produced inconsistent results. Cullum et al. (1991) and Reitan & Wolfson (1995, 1996, 1997) demonstrated the utility of examining the consistency of responses across test occasions by showing that litigants were far less consistent than non-litigants in neuropyschological test performance. These studies did not include a malingering condition that makes it difficult to draw conclusions from this study regarding malingerers versus bona fide patients.

Demakis (1999) used an analog study to examine test-retest effects and response consistency. The California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987), the Rey Complex Figure Test (RCF; Meyers & Meyers, 1995), the Controlled Oral Word Association Test (COWAT; Ruff, Light, Parker & Levin, 1996), and the Ruff Figural Fluency Test (RFFT; Ruff, 1988) were evaluated. The CVLT, a list learning task was chosen because it is a commonly used neuropsychological test that has been proven as a useful measure to detect malingering (Millis, Putnam, Adams, & Ricker, 1995). The other tests were chosen because they too are commonly used tests in forensic neuropsychology but little research has assessed malingering performance on them. Half of the participants were instructed to malinger on two sessions that were separated by a 3-week interval, and the other half were instructed to perform their best. The results indicated that malingering simulators performed less consistently than controls, but only when consistency was assessed via qualitative levels of performance

(percent recall consistency across the 5 trials in the CVLT). Contrary to prior studies, when consistency was assessed via quantitative levels of performance (variables that assess the number of words recalled), the two groups were equally consistent. Furthermore, Demakis (1999) also found a difference in the pattern of improvement between the verbal measures (in the CVLT and the COWAT), and the nonverbal measures (in the RCF and RFFT tests). The differential pattern of improvement on verbal and nonverbal measures, as well as the difference of consistency measures when considered quantitatively and qualitatively suggest the need for numerous tests, analyzed quantitatively and qualitatively in order to examine variability in performance.

One shortcoming of Demakis (1999)'s study, typically associated with analog studies, was the lack of bona fide patients to examine the potential impact of severity and location of head injury experienced, premorbid intelligence, and patient history. For example, variability in performance can potentially be symptomatic of specific head injury location (Bleiberg, Garmoe, Halpern, Reeves, & Nadler, 1997; Stuss, Murphy, & Binns, 1999). If this is the case, then inconsistency in performance would not be an indicator of malingering per se but also of neurological (e.g., frontal lobe) damage (Stuss et al., 1999). This would make it difficult to distinguish between malingerers and TBI patients because they could both have the same variability in performance, but for different reasons. The malingerers may not be able to feign their performance consistently across occasions while the TBI patients would merely be exhibiting a symptom of their specific injury.

Strauss et al. (1999) examined the performance variability of university students in an analog study of malingering. Using a symptom validity task (VSVT; Slick et al.,

1997) and traditional neuropsychological measures (Digit Span from the WAIS-III; Wechsler, 1997; 4 and 5 choice recognition vocabulary test; Ekstrom, French, Harman, & Dermen, 1976) across three separate occasions, this study demonstrated the utility of examining consistency across occasions as an indicator of malingering. They found variability in performance on the VSVT task, in both accuracy and reaction time measures. Furthermore, through logistic regression, it was revealed that across occasion variability in VSVT accuracy scores contributed uniquely to prediction accuracy both for hard and easy items. Predictive accuracy improved on hard items from 88.7% (first-occasion level of performance) to 96.1%, and variability on easy items improved predictive accuracy from 88.7% to 95.6%. Although it is an analog study, and university students could potentially behave differently from individuals who have suffered a head injury, it does warrant further investigation into the use of variability in performance as an indicator of malingering behaviour.

These results suggest that variability in performance could be a useful indicator of malingering. It is important, however, to consider a clinical non-litigating population, and examine their genuine variability in performance and compare that to those who may be malingerers. It is conceivable that both head-injured and malingerers show increased variability on cognitive tasks, however the nature of the variability may differ between groups. In order to examine the potential differences in variability across groups, it would be helpful to disassociate variability into the aforementioned concepts of intraindividual change and intraindividual variability. This leads to the question of whether variability between groups differs as a function of learning, as well as the

question of whether variability between groups differs as a function of some sort of central nervous system compromise, that could be the result of a head injury.

In the present study, we used a modified analog paradigm not commonly found in the literature. Individuals with expertise (namely professionals with knowledge of head injury and individuals with bona fide TBI) were contrasted with naïve participants. Within each of the three groups, half were asked to simulate a head injury in order to receive compensation and the other half were asked to perform their best on the tasks that followed. The design offers greater ecological validity due to the breadth of knowledge among our professionals, and the level of knowledge that exists with participants who have suffered a head injury. In particular we wished to examine whether knowledge makes a significant contribution to feigned performance.

There are several potential outcomes that could occur. First, knowledge could be a significant contributor to malingering, and as a result the head injured and experienced group may prove to be better malingerers than the naïve group. Another possibility is that type of knowledge differs across groups, with either the head-injured group malingering more successfully than the professionals, or the reverse. The third possibility is that the validity of the tests is robust to effects of knowledge, and no group effect would occur. These participants performed the tasks on three separate occasions in order to assess variability across occasions.

Several types of tasks (traditional neuropsychological tests, symptom validity tests) were used. In this way we could assess the utility of variability as an indicator of malingering and whether such variability varies as a function of tasks. Is variability in performance task specific thereby rendering some tests more useful then others as an

indicator of malingering? Furthermore, data were analyzed in such a way so as to examine both 'intraindividual change' as well as 'intraindividual variability' independent of mean performance.

Chapter II

METHOD

Participants

The sample consisted of 74 participants (27 head injured adults, 26 naïve people and 21 experienced individuals) recruited via advertisements and newspaper ads. The participants ranged in age from 18 to 57 years ($M = 36.24$, $SD = 9.68$). The first group consisted of 27 individuals diagnosed by their physician as having suffered a head injury. Nineteen individuals had suffered a mild head injury, 1 individual had a moderate head injury, and 7 individuals had suffered a severe head injury. The time lapsed since injury ranged from 1 to 28 years. Participants were excluded if they had a significant psychological disorder, a history of drug abuse, or were currently involved in litigation.

The second group consisted of 26 naïve individuals. The third group consisted of 21 experienced individuals. All experienced individuals had professional contact with individuals with head injuries, and could be broadly classified into one of the four groups: nurses, social service workers, rehabilitationists, and legal professionals. Exclusionary criteria for these latter two groups included the presence of any diseases that affect central nervous system functioning such as epilepsy or a heart condition. Other exclusionary criteria included the presence of significant health-related (including emotional) disorders, and the use of medications that might alter cognitive status.

Participants provided demographic and self-reported health information during an initial intake interview. In addition, several benchmark cognitive measures were obtained during this initial session including the MMSE (Folstein, Folstein, & McHugh, 1975), the Wechsler Adult Intelligence Scale-III (WAIS-III) Block Design, and Vocabulary subtests (Psychological Corporation, 1997), and the North American Adult Reading Test (NAART; Blair & Spreen, 1989). We also computed estimates of full-scale IQ (FSIQ) based on the age-adjusted Block Design and Vocabulary subtests (Sattler & Ryan, 1999), and premorbid IQ based on the NAART (Blair & Spreen, 1989).[1]

Table 1 shows the age, education, self-reported health, and benchmark cognitive status of the participants as a function of group.

---

[1] Blair and Spreen's (1989) formula for estimating premorbid intellectual ability using the NAART is : $NAART_{estIQ} = 127.8 - .78\ (NAART_{errors})$. This formula is based on the WAIS-R, whereas our estimate of current IQ is based on the WAIS-III which is somewhat more difficult. Thus, the discrepancy between the NAART estimate and the WAIS-III estimate will, if anything, slightly overestimate cognitive decline.

Table 1. Demographic and performance characteristics as a function of group

Variable		Group		
		Head Injured	Naive	Experienced
Age	M	36.15	32.77	40.67
	SD	8.70	10.01	9.06
Years of education	M	13.93	14.81	17.19
	SD	2.06	2.21	2.66
Health	M	4.41	4.27	4.43
	SD	0.80	0.67	0.51
MMSE	M	29.03	29.42	29.57
	SD	1.34	0.76	0.68
WAIS Block Design	M	13.41	11.81	11.48
	SD	2.95	2.77	2.99
WAIS Vocabulary	M	13.26	12.69	13.91
	SD	2.55	2.92	2.47
Estimated WAIS FSIQ	M	119.26	112.96	115.48
	SD	14.12	14.45	12.64
Estimated NAART IQ	M	111.65	107.76	113.17
	SD	7.40	9.64	7.41

Significant group differences were observed for age, $F(2, 71) = 4.21$, $p < .05$, $\eta^2 = .106$, education, $F(2, 71) = 11.58$, $p < .00$, $\eta^2 = .074$ and Block Design scaled score, $F(2, 71) = 3.19$, $p < .05$, $\eta^2 = .082$. Posthoc contrasts using Tukey's HSD ($p < .05$) indicated that

the effect for age was the result of the naïve participants being significantly younger than the experienced individuals. The difference in education was the result of the experienced group having a significantly greater number of years of schooling than the head injury and naive groups. The head-injured individuals obtained higher scores on the Block Design subtest than the other two groups. The differences between the groups in age, education and spatial ability are addressed further in the Results section.

Design and Procedure

The study took place across four occasions, consisting of an initial interview and three testing sessions either in the university or in the participant's home. On the first occasion demographic information was collected. On occasions two to four participants were repeatedly tested on the battery of malingering measures.

In the initial interview, participants were asked to answer the questions read to them from the demographic questionnaire. Afterward, the MMSE was administered to screen for cognitive impairment. Following the MMSE, baseline cognitive measures were taken. The two subtests of the WAIS (Block design and vocabulary) were given to obtain a rough estimate of IQ. An estimate of premorbid intellectual ability was obtained by administering the NAART.

After the initial interview, participants were tested individually over three measurement occasions scheduled approximately 2 weeks apart. Because this study was run concurrently with another study, the head-injured and naïve groups received an additional hour of cognitive testing prior to the tests involved in this experiment.

Prior to administering the malingering tests, participants were assigned to one of two experimental conditions. Individuals who were not malingering were simply told to

try their best. The simulators (who were randomly assigned to this group from the sample) were told that the purpose of this study was to develop accurate ways of discriminating exaggeration from true impairment. They were told that their motive was to seek extra compensation for their head injury, and in doing so, they were required to exaggerate their cognitive complaints. The following scenario was presented to participants in the malingering condition:

"A man ran a red light and hit the car you were driving. You injured your head in the accident. Now you have problems thinking and remembering things. The other driver is refusing to pay for injuring you, so you are suing him. The amount of money you will receive depends on how badly you were injured. You will try to get extra money by exaggerating your mental problems on the tests you are about to take. You must be careful because if you get caught exaggerating on the tests, you will not get any money from the man who caused your injury".

VSVT and Digit Span were given on each of the three measurement occasions. At the end of each session, the simulators were given an additional questionnaire assessing their comprehension of the scenario and their understanding of the role they were supposed to play. After the final session, the participants were all given a scale that was aimed at evaluating how hard they tried to play their given role.

Measures

The 21-item test (Iverson, Franzen & McCracken, 1991) is composed of a target list of 21 nouns that are read to the subject. Following the presentation of the list, the participant is instructed to recall as many words as they can. A second list of 21 nouns is added to the target list and the 21 pairs (the target word, and a foil) are read to the

participant. They are instructed to select the word that had previously been presented from the target list.

Digit Span was given under standard Wechsler Adult Intelligence Scale – III (WAIS-III; Wechsler, 1997) administration guidelines. Reliable Digit Span (RDS) was calculated by summing the longest string of digits repeated without error over two trials under both forward and backward conditions (Greiffenstein et al., 1996).

The Victoria Symptom Validity Test (VSVT; Slick et al., 1997) is a computerized forced-choice digit recognition task that evaluates both response accuracy and latency for 24 'easy' items (i.e., where items and foils share no digits in common such as 34092 and 56187), as well as 24 'hard' items (i.e., where items and foils are identical to the study number with the exception of a transposition of middle digits such as 46923 and 46293). Participants see a group of five digits on the screen for 5s followed either 5, 10 or 15s later by the target group of digits paired with a new group of digits. Participants are asked to choose as rapidly and accurately as possible the target group.

Screening and Manipulation Checks

In order to screen for suboptimal effort, all participants were given the 21-item test (Iverson, Franzen & McCracken, 1991). The two groups (malingerers, controls) did not differ in their performance on the forced choice component [$F(2, 70) = 0.60$, ns].[2] Participants in all groups scored 13 or greater on the task. Two severely head-injured individuals (assigned to the control condition) obtained scores of 13. All of the analyses were repeated excluding these two individuals, and the results were replicated.

At the end of each session, the malingerers were asked to fill out a questionnaire assessing how well they complied with the instructions to 'feign impairment after a car

accident'. Participants scores ranged from 3 to 6 out of 6 (M = 5.58, SD = 0.44). In addition, both malingerers and controls were asked to rate how hard they tried to fulfill the role they were asked to play on a 4-point scale (4 corresponding to high effort). A 3 (Group) X 2 (Malingering Instructions) ANOVA revealed a significant main effect of group [$F(2,67)=4.78$, $p<.01$]. Experienced individuals reported exerting more effort than the other groups. There were no other significant main effects. An a priori decision was made to exclude anyone who rated their performance at 1. No participants were excluded from the analyses. Scores ranged from 2 to 4 ($\underline{M}$ = 3.38, $\underline{SD}$ = 0.60).

CHAPTER III

Results

The results are presented in four main parts. First, level of performance was compared on the VSVT (latency and accuracy) and reliable digit span across groups, malingering instructions, and occasions. Next, for clinical purposes, the first occasion of measurement was examined on the VSVT and reliable digit span across groups and malingering instructions. This was done in order to discern the clinical utility of the tasks to evaluate first occasion level performance independent of variability analyses. The means and standard deviations for all measures across all three occasions are presented in Table 2. The third set of analyses examined group differences in intraindividual variability across the three occasions for both VSVT and digit span. Lastly, discriminant function analysis was used to examine how well the malingerers and controls could be differentiated on the basis of level of performance and intraindividual variability. We were particularly interested in whether intraindividual variability predicted group membership over and above level of performance.

---

[2] Data were missing for three individuals in the 'experienced' group.

Table 2

Means and Standard Deviations on all Variables for Sessions 1-3

		Variables				
		RDS	VSVT	VSVT	VSVT	VSVT
			Acc - easy	Acc - hard	RT - easy	RT - hard
Occasion 1	Control	11.05	23.88	22.87	1.53	2.16
		2.55	0.52	1.96	0.47	0.65
	Malinger	6.03	21.74	12.44	2.39	3.49
		3.42	3.21	6.45	1.15	1.49
Occasion 2	Control	11.38	23.80	22.68	1.55	2.18
		2.48	0.82	2.37	0.58	0.62
	Malinger	5.68	20.32	11.26	2.27	3.20
		2.36	4.17	5.45	1.02	1.75
Occasion 3	Control	10.83	23.80	22.85	1.55	2.06
		1.99	0.56	1.59	0.51	0.54
	Malinger	5.50	19.76	11.65	2.25	2.78
		2.95	4.45	5.87	1.18	1.11

Level of Performance

For the VSVT analyses, we computed separate 3 (Group) X 2 (Malingering Instructions) X 2 (Difficulty) X 3 (Occasion) repeated measures ANOVAS on the latency and accuracy measures, with Group and Malingering Instructions as between subjects factors, and difficulty and occasion varying within subjects. Similarly, separate 3 (Group) X 2 (Instructions) X 3 (Occasion) analyses were performed on the reliable digits measure.

As noted previously, there were significant group differences in age and education. Therefore all analyses were repeated covarying age, education and block design.

Level of performance on VSVT latency. The analysis of the VSVT latency scores indicated that there were significant main effects associated with Malingering Instructions, F (1, 68) = 26.55, $p < .001$, $\eta^2 = .28$, and Difficulty, F (1, 68) = 100.42, $p < .001$, $\eta^2 = .60$. Participants instructed to malinger were slower (M = 2.73, SD = 1.38) than participants who were instructed to 'do their best' (M = 1.84, SD = 0.63). In addition, all participants performed slower on hard items (M = 2.60, SD = 1.21) compared to easy items (M = 1.89, SD = 0.92). There was also a significant Difficulty X Occasion interaction, F (1, 68) = 5.43, $p < .01$, $\eta^2 = .74$. Simple effects contrasts were computed across difficulty within each occasion and difficulty was still found to be significant for each of the three occasions, despite the interaction that indicates that the discrepancy between easy and difficult items declines across occasions. However, the Difficulty X Occasion interaction for latency scores became insignificant after covarying for age and education.

Level of performance in VSVT accuracy. The analysis of the VSVT accuracy scores indicated that there were significant main effects associated with Instructions, F (1, 68) = 175.89, $p < .001$, $\eta^2 = .72$, and Difficulty, F (1, 68) = 147.20, $p < .001$, $\eta^2 = .68$, as well as an Instructions X Difficulty interaction, F (1, 68) = 93.64, $p < .001$, $\eta^2 = .58$.

As expected, malingerers were less accurate than controls on both easy and hard items, and participants instructed to malinger performed worse (M = 16.20, SD = 6.70) than participants instructed to do their best (M = 23.31, SD = 1.56). In addition, all participants performed better on the easy items (M = 22.35, SD = 3.20) than on the hard items (M = 17.74, SD = 6.95). These main effects were conditional on the interaction shown in Figure 1 which indicates that controls performed similarly on the easy and hard items, whereas malingerers performed substantially worse on the hard compared to the easy items.

The analysis was repeated covarying age, education and Block Design and identical results were found. The main effects of Instructions and Difficulty as well as the interaction remained significant.

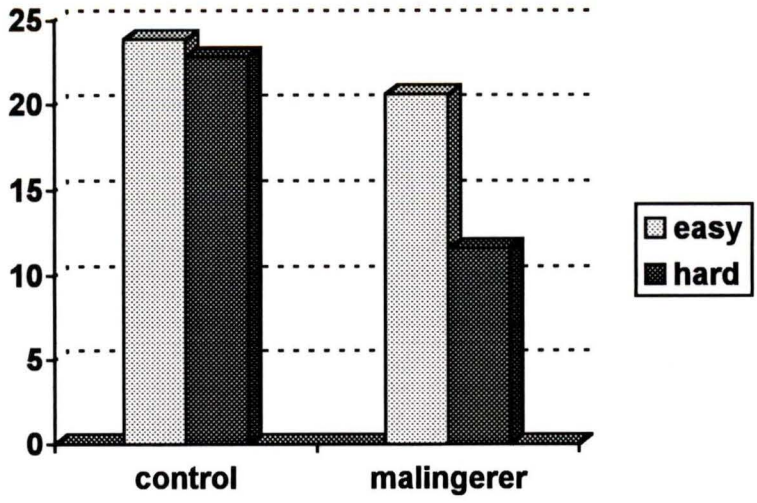


Figure 1. Performance on VSVT easy and hard (accuracy) as a function of malingering instructions.

Level of performance in Reliable Digit Span. Instructions had a significant effect on RDS, F (1, 67) = 124.604, $p < .001$, $\eta^2 = 647$. Malingerers performed worse (M = 5.74, SD = 2.92) than the controls (M = 11.08, SD = 2.34) on the reliable digit span. The analyses on the digit span scores were repeated covarying age, education and block design yielding the same results.

Level of Performance on Occasion One

The next focus was on the first occasion of measurement. All ANOVAs were repeated without the occasion effect. The analyses were than run covarying age and education and yielded similar results. The means and standard deviations for all measures are presented in Table 2.

Level of performance on VSVT latency. Main effects associated with Instructions, F (1, 68) = 28.75, $p<.001$, $\eta^2 = .29$, and Difficulty, F (1, 68) = 58.39, $p<.001$, $\eta^2 = .46$ had a significant effect on the VSVT latency scores. Controls performed faster than malingerers and response latencies were shorter for easy than for hard items. The Instructions X Difficulty interaction was also significant, F (2, 67) = 4.00, $p<.05$, $\eta^2 =$

.06, indicating that the discrepancy between groups (control, malingering) was most pronounced for the malingerers.

Level of performance on VSVT accuracy. The analysis revealed significant main effects for Instructions, F (1, 68) = 91.78, $p<.001$, $\eta^2 = .57$, and Difficulty, F (1, 68) = 97.44, $p<.001$, $\eta^2 = .59$. Regardless of experience, the control group was more accurate than the malingering group and accuracy was higher for easy than for hard items. The Instructions X Difficulty interaction was significant, F (2, 68) = 64.56, $p<.001$, $\eta^2 = .49$. The discrepancy between Instructions (malingering, control) was most pronounced for the hard items. Control and malingerers performed similarly on the easy items, but on the hard items, malingerers depressed their performance.

Level of performance on Reliable Digit Span. Regardless of prior knowledge (i.e. no group effect), malingerers performed worse than the controls on the RDS, F (2, 71) = 52.21, $p<.001$, $\eta^2 = .42$.

Intraindividual Variability

In order to examine intraindividual variability, our first concern was to purify the data for any systematic effects associated with occasion (e.g., practice effects), mean differences across groups, or level of difficulty (in VSVT data). In purifying for systematic effects, we can than analyze the inconsistency in performance independent of any durable change that is related to development or learning. The focus rather becomes the examination of intraindividual variability or cyclical fluctuations related to state and not trait characteristics. The purification procedure (consisting of regressing occasion, group, and level of difficulty on the trial scores) helps to dissociate state from trait variability allowing a statistical examination of the state variability in the given data set.

The residual scores (which many would consider error) can be considered as an estimate of intraindividual variability uncontaminated by systematic variation due to materials or practice effects. We then standardized the residuals by creating T scores in order to allow for comparison across the VSVT and Digit Span tasks. Next, we computed intraindividual standard deviations (ISD) from the residual scores for each individual across the three occasions. We also computed the coefficient of variation (CV) in which an individual's ISD is divided by their own mean score. This provides a measure of intraindividual variability relative to the individual's own level of performance.

For these analyses, we used univariate ANOVAs in order to examine the effects of Group, Instructions, and Difficulty for the VSVT and Group and Instructions for reliable digit span. We also replicated all analyses covarying age and education. Correlations between age and education and the variables ranged from -.009 to .237. The correlation between education and VSVT accuracy on hard items was significant (r = .237, $p < .05$).

Intraindividual variability in VSVT latency across occasions. We computed a 3 (Group) X 2 (Instructions) X 2 (Difficulty) repeated measures ANOVA on the across occasions latency ISDs and CVs. The analysis of the ISDs revealed there was a significant main effect associated with Instructions, F (1, 68) = 17.21, p <.001, η^2 =.20 indicating that malingerers were more variable in their reaction times (M = 6.83, SD = 6.93) than controls (M = 2.74, SD =2.02). There was also a Group X Difficulty interaction approaching significance, F (2, 68) = 2.93, p <.060, η^2 =.079. Simple effects contrasts were computed across groups within each level of difficulty and difficulty was found to be significant only for the naïve group ($p<.05$). In the naïve group, level of

difficulty determined the extent of variability in performance. This effect was only marginally significant and disappeared after covarying for age, education and block design.

When looking at the same analyses using CVs, the Difficulty effect approached significance, $\underline{F}$ (1, 68) = 3.86, $\underline{p}$ <.054, η^2 =.054. Both groups were more variable on the hard items ($\underline{M}$ = 0.10, $\underline{SD}$ = 0.09) than on the easy items ($\underline{M}$ = 0.08, $\underline{SD}$ = 0.08). In short, Instructions had an effect with both ISD and CV.

After covarying for age, education and Block Design, the Difficulty effect and the Difficulty X Group interaction were no longer significant ($\underline{p}$ <.16 and $\underline{p}$ <.31 respectively). The Instructions effect remained significant for both ISD and CV. Despite group differences in age and education, Instructional set impacts level of variability on VSVT latency scores.

Intraindividual variability in VSVT accuracy across occasions. We computed a 3 (Group) X 2 (Instructions) X 2 (Difficulty) repeated measures ANOVA on the across occasions accuracy ISDs and CVs. The analysis of ISDs showed a significant main effect associated with Instructions, $\underline{F}$ (1, 68) = 22.59, $\underline{p}$ <.001, η^2 =.25. The malingerers showed more variability in the accuracy of their responses ($\underline{M}$ = 6.14, $\underline{SD}$ = 6.34) than controls ($\underline{M}$ = 1.87, $\underline{SD}$ = 1.82) on both the easy and hard items combined. The significant Difficulty effect, $\underline{F}$ (1, 68) = 16.61, $\underline{p}$ <.001, η^2 =.20 indicated that all groups performed more variably on hard items ($\underline{M}$ = 4.97, $\underline{SD}$ = 5.63) than on easy items ($\underline{M}$ = 2.71, $\underline{SD}$ = 3.92). There was no significant interaction between Instructions and Difficulty.

When looking at the same analyses using CVs, the Difficulty effect and Instructions effect remained significant at the same probability levels as the analyses using ISDs. After covarying for age and education, the results were identical for ISDs and for CVs with the exception of one interaction approaching significance (CV). There was a Malingering Instructions X Difficulty interaction, $\underline{F}$ (1, 66) = 3.22, $\underline{p}$ <.077, η^2 =.047 that approached significance. Malingerers performed more variably on hard items (M = 0.171, SD = 0.157) than on easy items ($\underline{M}$ = 0.096, $\underline{SD}$ = 0.119). Controls seemed to have less of a discrepancy between variability on easy items ($\underline{M}$ = 0.022, $\underline{SD}$ = 0.013) and variability on hard items ($\underline{M}$ = 0.061, $\underline{SD}$ = 0.071).

Intraindividual variability in Reliable Digit Span. We computed a 3 (Group) X 2 (Malingering Instructions) ANOVA on across occasion reliable digit span ISDs and CVs. There were significant main effects associated with Instructions, $\underline{F}$ (1, 68) = 13.14, $\underline{p}$ <.001, η^2 =.162. The malingerers showed more variability across the 3 occasions ($\underline{M}$ = 7.75, $\underline{SD}$ = 5.41) than the controls ($\underline{M}$ = 4.44, $\underline{SD}$ = 2.81). Analyzing CVs produced the same effect associated with Instructions. In addition, after covarying for age, education and Block Design the identical results were found.

Predicting Group Membership

Previous analyses indicated that the malingering and control groups differed in level of performance as well as in intraindividual variability. There was no indication that performance or variability differentiated between groups of naïve, head injured, or experienced participants. Consequently, the next question addressed how well the combined performance and variability information differentiated malingerers from the controls, collapsed across all 3 groups of participants. Of particular interest is whether

performance and variability make independent contributions to predicting group membership. Thus, discriminant function analysis was used to estimate the extent to which malingerers and controls (combined) could be differentiated, as well as to evaluate the unique contribution of sets of level of performance variables versus variability variables. For the VSVT, the variables included in the discriminant function analysis were accuracy, latency, across occasions latency ISD, and across occasions accuracy ISD. These variables were entered for both hard and easy items. For Reliable Digits, total reliable digit span and an across occasions reliable digit span ISD were used. The performance variables were taken from occasion one given that there was no occasion effect revealed in the previous analyses, and using the first occasion variables is more relevant in many clinical contexts.

First, the discriminant function for the performance variables was evaluated. Next, we looked at the predictability of the variability measures on their own. The last discriminant function included both the performance and variability measures in one classification analysis to examine predictability extracting all of the information from the three occasions. Finally, in order to evaluate the unique contribution of both the performance and variability measures, the models are compared to one another using hierarchical regression.

Discriminant function on the performance measures. The results for both VSVT and Digit Span indicated that the performance information differentiated the malingerers from the non-malingerers, χ^2 (5) = 68.40, $p < .00$, Wilks' Lambda = .374, correctly classifying 97.50% of non-malingerers, 79.40% of malingerers, and 89.20% overall. In looking at the performance variables individually, VSVT accuracy for hard items is the

most significant predictor of group membership, t (68) = -4.67, $p < .001$. Reliable Digit Span accuracy also contributes uniquely to predicting group membership, t (68) = -3.12, $p < .001$. The rest of the performance variables did not provide a significantly unique contribution to predicting group membership.

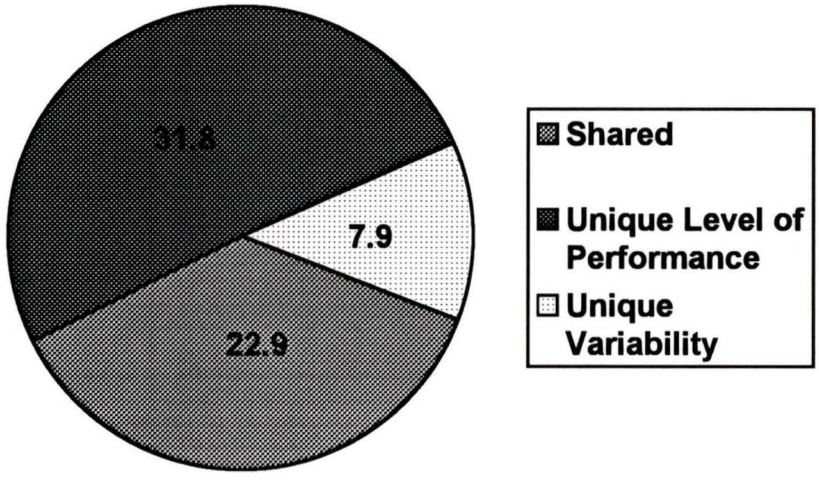
Discriminant function on the variability measures. The analyses differentiated the malingerers from the non-malingerers, χ^2 (5) = 34.01, $p < .00$, Wilks' Lambda = .613, correctly classifying 97.50% of the non-malingerers, 58.80% of the malingerers, and 79.70% overall. The ISD for VSVT latency on hard items made the most unique contribution to group membership, t (68) = 2.42, $p < .05$. In addition, the ISD for VSVT accuracy on hard items approached significance at t (68) = 1.94, $p < .057$. The reliable digits ISD measure made a significantly unique contribution to group membership as well, t (68) = 2.50, $p < .015$, increasing both sensitivity and specificity of the classification.

Discriminant function on performance and variability measures. Performance and variability predictors differentiated the malingerers from the non-malingerers, χ^2 (10) = 81.74, $p < .001$, Wilks' Lambda = .295, correctly classifying 97.50% of the non-malingerers, 85.30% of the malingerers, and 91.90% overall. When combining performance and variability, VSVT accuracy for hard items was the only predictor that significantly contributes to group classification uniquely, t (63) = -4.07, $p < .001$. Reliable Digit Span approached significance, t (63) = -1.90, $p < .063$ as a unique contributor to group classification.

The question of the variability measures contributing to group classification over and above the performance measures was addressed through comparing models of

performance to that of performance and variability. Variability measures accounted for 7.90% of the variance over and above performance variables. The change in predictability was significant, F (5, 63) = 3.35, $p < .01$. When comparing models of variability to that of variability and performance, performance accounted for 31.80% of the variance over and above that of variability. The change in predicting group membership was significant, F (5, 63) = 13.56, $p < .001$. Both sets of predictors make a unique and significant contribution to predicting group membership as shown in Figure 4 below.

Figure 4. The unique contributions of variability and level of performance



CHAPTER IV

Discussion

The aim of this study was to evaluate the utility of various measures of malingering using an analog design in which half the participants were asked to try their best and the remainder were asked to feign believable injury. To do this, three groups

were examined: naïve healthy participants, professionals with knowledge of head injury, and individuals with documented evidence of a traumatic brain who were not in litigation at the time of study. In this way, we could evaluate the potential contribution of expert knowledge /experience as well as neurological insult to test performance. Further, participants were assessed on three separate occasions in order to assess whether repeat administration of the tests, especially the forced-choice measure of malingering, improves prediction.

Significance of Malingering Measures on a Single Test Occasion

Consistent with previous research (e.g., Greiffenstein et al., 1994; Rose et al., 1998; Strauss et al., 1999), our findings suggest that on a single test session, forced-choice recognition tasks achieve an excellent hit rate for the detection of noncompliance. The VSVT Hard Accuracy score loaded highest on the structure matrix indicating that it is the best predictor amongst the performance variables (Reliable Digit Span and the other VSVT variables). Slick et al. (1997) have recommended that, for hard items, correct scores above chance (that is, 16 or more) represent a valid score whereas a score below chance (8 or less) is unequivocally invalid. Scores within the chance interval (9-15) are classified as questionable. Using this guideline, 39 of 40 control participants (97.50%) performed above chance. One individual scored within the chance interval. By contrast, only 26.50% of those in the malingering group (9/34) successfully feigned impairment by performing in the above chance range. There were 19 out of 34 individuals scoring within the questionable range, and 17.65% (6/34) scoring in the invalid range. A similar pattern emerged when only the head-injured individuals were considered. This implies that the results of this study could be generalized to a patient population who have

suffered closed head injuries, and that no differences in the current study were found as a result of some of the participants having been concussed.

The Reliable Digit Span was the next predictor that loaded high on the structure matrix. Using the recommended cutoff score of <7 (Greiffenstein, Gola, Baker, 1994; Greiffenstein et al., 1995), most of the controls (38/40) were correctly classified. However, sensitivity was poor, correctly classifying only 52.94% of the malingerers.

Both the Reliable Digit Span and the VSVT scores contribute significantly to predicting group membership, correctly classifying 97.50% of the controls and 79.40% of the malingerers when entered simultaneously in a discriminant function analysis. Entered into the analysis individually, the Reliable Digit Span correctly classified 90.00% of controls and 79.40% of malingerers; the VSVT correctly classified 97.50% of controls and 76.50% of malingerers. This is illustrative of the potential clinical utility in using both measures, to increase both the specificity and sensitivity of the diagnostic battery. Additional research is necessary to determine whether use of other forced-choice tasks (e.g., Word Memory Test, TOMM) in conjunction with the VSVT would provide additional significant clinical benefit.

Significance of Variability in Performance Across Three Test Occasions

In line with other findings (e.g., Cullum et al., 1991; Demakis, 1999; Reitan & Wolfson, 1995, 1996, 1997; Strauss et al., 1999), our results supported the notion that examination of consistency of responses across test occasions is useful. More specifically, we found significant group differences in variability for both the Reliable Digit Span and the VSVT. Strauss et al. (1999) found only significant group differences in variability on the VSVT measures. Additionally, while Demakis (1999) did find the

examination of variability as a significant contributor to classifying malingerers, this finding was dependent on the type of tasks that were given. That is, variability could be task dependent. Although this study supports the notion that variability in performance is a significant indicator of malingering, it also emphasizes the need to determine the possibility of task dependence in variability measures.

Furthermore, in Strauss et al.(1999), both variability measures of VSVT latency and accuracy yielded a Group (malingerers and controls) by Item Difficulty interaction on session one performance data. In this study, session one analyses yielded this interaction but averaging across all 3 sessions, this interaction was found to be insignificant. Additionally, in this study, the difficulty effect disappeared for the latency score after covarying for group differences in age and education. These analyses were rerun with just the naïve and professional individuals in attempt to replicate the previous study but the results did not change significantly from those reported in this paper. There are at least two possible reasons for the potentially discrepant finding. First, the difference in demographics in the study samples could have had an impact on performance. That is, university students could differ in their conceptions of malingering and performance on cognitive tests from older participants that are in the workforce. Second, two-thirds of the current sample received a large battery of cognitive tests before the malingering tests and this could have impacted their performance. In the current study, successful feigning of the same levels of effort across different test sessions appears to be difficult for malingerers.

Variability in performance versus level of performance. One could argue that the group differences in variability are inevitable due to the high correlation between

individual performance and variability. That is, if performance differentiates groups than variability must logically follow. The results did also show that variability contributed uniquely to group classification over and above performance variables. Hierarchical regression demonstrates that variability accounts for 7.90% of the variance over and above performance scores. Although this is significantly different from the 31.80% of unique contribution that the performance variables hold, it still makes a significant and unique contribution to classifying malingering. The implication is that in order to increase sensitivity and specificity rates in classifying malingerers, one could look at variability as a unique indicator in addition to the given diagnostic battery. The utility of variability as an indicator of malingering has the additional possibility of being robust to coaching effects. From a clinical perspective, variability does not add significantly to warrant its usage in addition to the high classification rates that were reported with both the VSVT and RDS.

The Effects of Coaching and/or Knowledge of TBI

A limitation of previous analog studies is that university students were asked to simulate brain injury. Such individuals may differ markedly from persons choosing to malinger in a medical-legal context. Persons involved in litigation following even a mild traumatic brain injury may have considerable knowledge of head injury symptoms through direct experience, exposure to other patients and inadvertent or deliberate coaching. They also typically have more experience with neuropsychological tests, more time to prepare for the assessment and are more motivated to perform in a manner consistent with a diagnosis of TBI (Rogers, 1997). In this study, the measures were not affected by past experience with head injury. That is, head-injured individuals,

professionals, and naïve individuals altered their response style similarly according to instructional set. A potential limitation with this study is the varied amount of knowledge amongst the groups. That is, some head-injured individuals had been involved in litigation and were involved with neuropsychological testing, whereas others had only sought medical treatment. Additionally, professionals varied in the type of contact they had with patients, as well as the number of years experience they had working with head-injured people. Documenting the level of knowledge that individuals have about concussed individuals could help in providing a more thorough conclusion about the contribution of knowledge to feigning injury. Furthermore, knowledge as defined by experience may have no relevance to malingering performance and knowledge about tests and the expected results could be more important. Although this study found that knowledge does not contribute to malingering status, coaching may still render various indices of malingering insensitive (e.g., Coleman et al., 1998; Rose et al., 1998).

Examining The Etiology and Significance of Variability in Performance

Given the apparent difficulty in developing indices that are robust to knowledge and coaching, the examination of variability becomes more important. It is more difficult to coach someone on the appropriate level of variability to exhibit than it is to tell someone to suppress their performance on a test that day. Given this fact, the investigation of variability in performance is of paramount importance. The utility of variability in predicting malingering is contingent upon a greater understanding of what causes variability in performance. If a central nervous system compromise is indicative of head injury, head-injured individuals could theoretically perform more variably than controls given the hypothesis that variability in performance is caused by CNS

compromise. With this potential, it would be difficult to assume the performance of malingerers as being more variable than controls without knowing the nature of the variability that could exist amongst bona fide patients. Our analyses were repeated with just the head-injured individuals and the same results were found. That is, any variability in performance that could have been a result of an individual's head injury did not seem to impact the results. The head injured individuals yielded the same results as the controls and professional. This could be due to variability among our concussed sample being task dependant (i.e. variability could be present in tasks that were not in this study) or alternatively the nature of the injuries, and the time since injury could minimize variability in performance that is typically expected with concussed individuals.

Summary

The present findings suggest that consideration of both level of performance and intraindividual variability hold considerable promise for the detection of malingering. We recognize that our findings require independent cross-validation and further research to determine the nature of variability that should be expected amongst head-injured given the severity and location of their injury. Additionally, it would be useful to have tasks that did elicit variable performance as well as tasks that were typically less variable. The comparison of these tasks on levels of variability would increase the lines of evidence that one could use in order to determine a diagnosis. We do emphasize that diagnosis of malingering cannot be made on the basis of a single index. Rather, a competent assessment requires an in-depth interview, thorough review of the entire record and neuropsychological profile and that alternative diagnoses be ruled out (e.g., Slick et al., 1999; Tombaugh, 1996).

References

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, D.C: Author.

Bernard, L.C. (1991). The detection of faked deficits on the Rey Auditory Verbal Learning Test: The effect of serial position. Archives of Clinical Neuropsychology, 6, 81-88.

Bernstein, D. M. (1999). Recovering from mild head injury. Brain Injury, 13, 151-172.

Bickart, W.T., Meyer, R.G., & Connell, D. (1991). The symptom validity technique as a measure of feigned short-term memory deficit. American Journal of Forensic Psychology, 9, 3-11.

Binder, L. M. (1990). Malingering following minor head trauma. The Clinical Neuropsychologist, 4, 25-36.

Binder, L.M., Villaneuva, M.R., Howieson, D., & Moore, R.T. (1993). The Rey AVLT Recognition Memory Task measures motivational impairment after mild head trauma. Archives of Clinical Neuropsychology, 8, 137-149.

Binder, L. M. (1997). A review of mild head trauma. Journal of Clinical and Experimental Neuropsychology, 19, 432-457.

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

Blair, J.R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test, The Clinical Neuropsychologist, 3, 129-136.

Bleiberg, J., Garmoe, W. S., Halpern, E. L., Reeves, D. L., & Nadler, J. D. (1997). Consistency of within-day and across-day performance after mild brain injury. Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 10, 247-253.

Cullum, C., Heaton, R., & Grant, I. (1991). Psychogenic factors influencing neuropsychological performance: Somatoform disorders, factitious disorders, and malingering. In H.O. Doerr & A. S. Carlin (Eds.), Forensic Neuropsychology: Legal and scientific bases (pp. 141-171). New York: Guilford Press.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, E.A. (1987). The California Verbal Learning Test: Research edition. New York: The Psychological Corporation.

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

Ekstrom, R.B., French, J.W., Harman, H.H., & Derman D. (1976). Manual for the kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). 'Mini-mental State'. A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12, 189-198.

Franzen, M.D., & Martin, N. (1996). Do people with knowledge fake better? Applied Neuropsychology, 3, 82-85.

Gouvier, W. D., Hayes, J. S., & Smiroldo, B. B. (1998). The significance of base rates, test sensitivity, test specificity, and subjects' knowledge of symptoms in assessing

TBI sequelae and malingering. In R. Reynolds (Ed.), Detection of malingering during head injury litigation. New York: Plenum Press.

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

Greiffenstein, M.F., Baker, W.J., & Gola, T. (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. The Clinical Neuropsychologist, 9, 230-240.

Greiffenstein, M.F., Baker, W.J., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and postconcussion syndrome. Journal of International Neuropsychological Society, 2, 477-485.

Hayes, J.S., Hilsabeck, R.C., Gouvier, W.D. (1999). Malingering traumatic brain injury: Current issues and caveats in assessment and classification. In Varney, N.R. & Roberts, R. (Eds.), The evaluation and treatment of mild traumatic brain injury. Mahwah, NJ: Lawrence Erlbaum Associates.

Heaton, R.K., Smith, H.H., Lehman, R. A, & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. Journal of Consulting and Clinical Psychology, 46, 892-900.

Heubrock, D., & Peterman, F. (1998). Neuropsychological assessment of suspected malingering: Research results, evaluation techniques, and further directions of research and application. European Journal of Psychological Assessment, 14, 211-225.

Iverson, G.L., Franzen, M. D., & McCracken, L. M. (1991). Application of a forced-choice memory procedure designed to detect experimental malingering. Archives of Clinical Neuropsychology, 15, 667-676.

Iverson, G.L., Franzen, M. D., & McCracken, L. M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. Law and Human Behaviour, 15, 667-676.

Iverson, G. L., & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. Journal of Clinical and Experimental Neuropsychology, 18, 38-51.

Kaplan, E.F., Goodglass, H., & Weintraub, S. (1983). The Boston Naming Test (2nd ed.). Philadelphia: Lea & Febiger.

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

Levin, H. S., Benton, H. L., & Grossman, R.G. (1982). Neurobehavioural consequences of closed head injury. New York: Oxford.

Martin, R. C., Franzen, M. D., & Orey, S. (1998). Magnitude of error as a strategy to detect feigned memory impairment. The Clinical Neuropsychologist, 12, 84-91.

Martin, R.C., Hayes, J..S., & Gouvier, W.D., Todd, M.E., Bolter, J.F., & Niccolls, R. (1992). Effects of task instruction on malingered memory performance. Forensic Reports, 5, 393-397.

Meyers, J.E., & Meyers, K.R. (1995). Rey Complex Figure Test and Recognition Trial: Professional manual. Odessa, FL: Psychological Assessment Resources.

Meyers, J.E., & Volbrecht, M. (1998). Validation of Reliable Digits for detection of malingering. Assessment, 5, 303-307.

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

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

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

Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In R. Downs, L. Liben, & D. S. Palerno (Eds.), Visions of aesthetics, the environment, and development: The legacy of Joachim F. Wohlwill (pp. 213-240). Hillsdale, NJ: Erlbaum.

Paul, D. S., Franzen, M. D., Cohen, S. H., & Fremouw, W. (1992). An investigation into the reliability and validity of two tests used in the detection of dissimulation. International Journal of Clinical Neuropsychology, 14, 1-9.

Psychological Corporation. (1997). WAIS-III Technical Manual. San Antonio, TX: Psychological Corporation.

Rapport, L. J., Frachoine, T. J., Coleman, R. D., & Axelrod, B. N. (1998). Effects of coaching on malingered motor function profiles. Journal of Clinical and Experimental Neuropsychology, 20, 89-97.

Reitan, R. M. (1958). Validity of the Trial Making Tests as an indicator of organic brain damage. Perceptual and Motor Skills, 8, 271-276.

Reitan, R. M., & Wolfson, D. (1995). Consistency of response on retesting among head-injured subjects in litigation versus head-injured subjects not in litigation. Applied Neuropsychology, 2, 67-71.

Reitan, R. M., & Wolfson, D. (1996). The question of validity of neuropsychological test scores among head-injured litigants: Development of a dissimulation index. Archives of Clinical Neuropsychology, 11, 573-580.

Reitan, R. M., & Wolfson, D. (1997). Consistency of neuropsychological test scores of head-injured subjects involved in litigation compared with head-injured subjects not involved in litigation: Development of the Retest Consistency Index. The Clinical Psychologist, 11, 69-76.

Rey, A. (1941). L'examen clinique en psychologie. (The clinical examination in psychology). Paris: Presses Universitaires de France.

Rey, A. (1941). L'examen psychologie dans las cas d'encephalopathie traumatique. Archives de Psychologie, 23(112), 286-340.

Rogers, R., Harrell, E.H., & Liff C.D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. Clinical Psychology Review, 13, 255-274.

Rose, F. E., Hall, S., & Szalda-Petree, A. D. (1998). A comparison of four tests of malingering and the effects of coaching. Archives of Clincial Neuropsychology, 13, 349-363.

Ruff, R. (1988). Ruff Figural Fluency Test administration manual. San Francisco, CA: Neuropsychological Resources.

Ruff, R.M., Light, R.H., Parker, S.B., & Levin H.S. (1996). Benton Controlled Oral Word Association Test: Reliability and updated norms. Archives of Clinical Neuropsychology, 11, 329-338.

Sattler, J.M., & Ryan, J.J. (1999). Assessment of children: Revised and Updated Third Edition WAIS-III Supplement. San Diego: Jerome M. Sattler.

Sewick, B. G., & Sobota, W. L. (1996). A cross validation study of the Reitan-Wolfson Dissimulation Index. Paper presented at the annual meeting of the American Psychological Association, Toronto, Canada.

Slick, D., Hopp, G., 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's Forced-Choice Memory Test. Journal of Clinical and Experimental Neuropsychology, 16, 472-481.

Slick, D. J., Hopp, G. H., Strauss, E., & Spellacy, F. J. (1996). Victoria symptom validity test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. Journal of Clinical and Experimental Neuropsychology, 18, 911-922.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. (1997). Victoria Symptom Validity Test Odessa, FL: Psychological Assessment Resources.

Spreen, O., & Strauss, E. (1991). A compendium of neuropsychological tests. New York: Oxford University Press.

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, 483-490.

Strauss, E, Hultsch, D. F., Hunter, M., Slick, D., Patry, B., & Levy-Bencheton, J. (1999). Using intraindividual variability to detect malingering in cognitive performance. The Clinical Neuropsychologist, 13, 420-432.

Stuss, D.T., Stethem, L.L., Hugenholtz, H., Picton, T., Pivik, J., & Richard, M.T. (1989). Reaction time after head injury: Fatigue, divided and focused attention, and consistency of performance. Journal of Neurology, Neurosurgery, and Psychiatry, 52, 742-748.

Stuss, D.T., Murphy, K.J., & Binns, M.A. (1999). The frontal lobes and performance variability: Evidence from reaction time. Journal of the international Neuropsyhological Society, 5, 123.

Stuss, D.T., Pogue, J., Buckle, L., & Bondar, J. (1994). Characterization of stability of performance in patients with traumatic brain injury: Variability and consistency on reaction time tests. Neuropsychology, 8, 316-324.

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

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

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 neuopsychological test identify malingering? A preliminary analysis. Journal of Clinical and Experimental Neuropsychology, 21, 245-250.

Wechsler, D. (1955). Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corporation.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale-Revised manual. New York: Psychological Corporation.

Wechsler, D. (1987). Wechsler Memory Scale- Revised. San Antonio, TX: The Psychological Corporation.

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

Wetter, M.W., & Corrigan, S.K. (1995). Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. Professional Psychology: Research and Practice, 26(5), 474-477.

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

VITA

Surname: Levy-Bencheton Given Name: Judith

Place of Birth: Toronto, Ontario

Education Institutions Attended:

York University	1994 to 1998
Hebrew University of Jerusalem	1995 to 1996
University of Victoria	1998 to present

Degrees Awarded:

B.A.	York University	1998

Honours and Awards:

Certificate of Excellent Academic Standing	1997, 1998

Publications:

Strauss, E, Hultsch, D. F., Hunter, M., Slick, D., Patry, B., & Levy-Bencheton, J. (1999). Using intraindividual variability to detect malingering in cognitive performance. The Clinical Neuropsychologist, 13, 420-432.

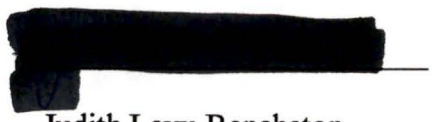
Hultsch, D.F., MacDonald, S.W., Hunter, M., Levy-Bencheton, J., & Strauss, E. (in press). Intraindividual variability in cognitive performance in the elderly: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. Neuropsychology.

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library or any other university, or similar institution on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

Validity of VSVT and Reliable Digit Span Measures for Detecting Malingering in Groups Differing In Exposure to and Knowledge of TBI.

Author 

Judith Levy-Bencheton

November __, 2000