THE DETECTION OF NEUROPSYCHOLOGICAL MALINGERING

Christine D. Liff, B.A., M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2003

APPROVED:

Craig Neumann, Acting Chair
Ernest Harrell, Chair, Major Professor, and
  Chair of the Department of Psychology
Kenneth Sewell, Committee Member and
  Director of Clinical Training
James Quinn, Committee Member
C. Neal Tate, Dean of the Robert B. Toulouse
  School of Graduate Studies

Liff, Christine D., *The Detection of Neuropsychological Malingering*. Doctor of Philosophy (Clinical Psychology), August 2003, 163 pp., 24 tables, 25 figures, references, 112 titles.

The present study compared the responses of a group of simulating malingerers who were offered a monetary incentive to feign symptoms of a head injury, with the responses of head injured groups both with and without litigation, a forensic parole group, and an honest-responding control group. The following six neuropsychological measures were utilized: Rey 15-Item Memory Test, Controlled Oral Word Association Test, Finger Oscillation Test, WAIS-R Neuropsychological Instrument (Vocabulary, Information, and Similarities subtests), Booklet Category Test, and Wisconsin Card Sorting Test. The statistical concepts of floor effect, performance curve, and magnitude of error were examined. Additionally, the statistical differences in the responses of the five groups were analyzed to determine cutting scores for use in distinguishing malingerers from nonmalingerers.

ACKNOWLEDGMENTS

To my grandmother, Elayne, who fostered my sense of adventure and passion for life. If only you could be here to share this with me.

To my parents, Leilani and George, who believed in hard work and allowed me autonomy and independence.

To my mentors along the way: Bruce, Neil, Shelly, Vi, Ann, Jim, Munro, Anne-Marie, and Pat. You, along with a handful of others, shaped me professionally, leaving me with more than I could ever repay.

Most of all, to my son, Evan, who has traveled this journey with me. He was my inspiration when things were difficult and resources lean. Through all the smiles and the tears, those we’ve shared our lives with, and those we’ve left behind, he is my constant, my shining star. I will forever be grateful for the gift his love has given me, and for being so blessed to share even a part of his life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Historically, the role of psychologists in litigation has focused upon the relationship between psychological factors and individual behavior. In recent years; however, this role has changed to include discussion of mental disorders that result from brain dysfunction or damage; disorders that are commonly studied by neuropsychologists (Youngjohn, 1991). One of the most common issues between litigation and neuropsychology is the determination and severity of brain dysfunction in personal injury, workers= compensation, and social security disability cases (Youngjohn, 1991). Neuropsychological test batteries, along with intelligence and personality testing, clinical interviews, and collateral information, are increasingly being utilized to assist with the diagnosis and severity of cognitive, emotional, and functional deficits due to head injury (Binder, 1990). In particular, the validity of an individual=s testing performance, and by association the validity of the injury, can come into question. The intentional distortion or exaggeration of testing performance, known as malingering, is increasingly encountered during litigation. As Gallucci (1984) states, there will likely be an increase of feigning in workmen=s compensation and personal injury claims which involve large financial incentives. The potential for malingering on the growing number of neuropsychological assessments, and the related litigation and financial burden are enormous. Financial settlements obtained through malingered neuropsychological deficits can be considerable, with claimants receiving unjust monetary awards (Franzen, Iverson & McCracken, 1990; Lee, Loring, & Martin, 1992), as well as medical costs and loss of earnings. This suggests the need for increased vigilance in the detection of malingering using neuropsychological instruments. Additional concerns regarding the need for the correct identification of malingerers involve both access to limited resources by a person who does not need them, or conversely, denying services to individuals falsely labeled as malingering (Franzen et al., 1990). Finally, the need for correct identification of malingerers is pressing as psychologists, and neuropsychologists in particular, are increasingly more involved in litigation (Resnick, 1988). As the field of neuropsychology continues to grow, so will its relationship to forensic issues (Larrabee, 1990). Yet surprisingly, even as neuropsychological malingering cases continue to increase, clinical studies on neuropsychological defensiveness and malingering remain limited, although this area was identified as one of the most important in malingering research (Rogers, 1984). As stated by Rogers, Harrell and Liff, Athe importance of assessing feigned cognitive deficits should not be underestimated@ (1993).

Neuropsychology, Malingering, and the Legal System

The classic definition of malingering -- the deliberate exaggeration or falsification of psychological and/or physical complaints, as a response to an external incentive, such as avoidance of negative circumstances or obtaining a tangible gain, was put forth by the American Psychiatric Association in 1994 (American Psychiatric Association; APA, 1994). Although the majority of civil and criminal court cases center around genuine neuropsychological issues, the feigning or exaggeration of neuropsychological deficits is not as uncommon as it was in the past (Youngjohn, 1991), with the number of cases involving suspected malingering continuing to rise (Resnick, 1988; Youngjohn, 1991). A literature review of malingering of psychological symptoms following physical injury cited incidence estimates ranging from 1 to 50% (Resnick, 1988), with 47% coming from workers= compensation claims (Marlowe, 1991, as cited in Youngjohn, 1991), and as high as 64% of individuals involved in personal injury claims (Heaton, Smith, Lehman, & Vogt, 1978). According to Gorman (1984), malingering is not a physical or psychological disorder; rather, it is the consequence of the individual=s volition, a legally wrongful act. Malingerers are often experienced deceivers who have much to lose if they are detected (Binder, 1992a). According to Rogers= (1990) adaptational model of malingering, there are three cognitive processes an individual engages in prior to malingering: A(a) a person perceives the evaluation/treatment as involuntary or adversarial, (b) the person perceives that [he/she] has...something to lose from self-disclosure or something to gain from malingering, and (c) the person does not perceive a more effective means to achieve [the] desired goal@ (p. 4).

Classifications and Definitions of Malingering

Classifications of malingering vary, as the problem of categorizing malingering remains an issue. A review of the literature; however, reveals several, often utilized classifications of malingering, ranging from pure malingering (an individual falsifies all of his/her symptoms) to partial malingering (an individual exaggerates the extent of his/her deficits) (Resnick, 1992). Rogers (1988a) identifies three general levels, or gradations of malingering: (a) mild malingering which involves primarily exaggeration, with minimal distortion and only a minor role in differential diagnosis, (b) moderate malingering which involves either exaggeration or fabrication by the individual to present his/her deficits as more significant than they actually are, and ranges from a small number to a variety of symptoms, and (c) severe malingering which involves such extreme fabrication of symptomatology that the presentation is absurd or incredible. Resnick (1992) further delineates malingering terms and definitions: (a) simulation, also called positive malingering, in which an individual feigns symptoms or a disorder that he or she does not possess, typically utilized for research purposes; and, (b) dissimulation, in which the individual conceals or minimizes symptoms or problems which would explain deficits (e.g. drug use) which classifies genuine malingering. Another malingering category identified by Resnick (1992) is false imputation in which the individual has actual deficits, but attributes them to something other than the cause (i.e. blaming an automobile accident when the injury actually occurred in another setting).

More recent classification reveals a set of diagnostic criteria outlined by Slick, Sherman, and Iverson, who address possible, probable, and definite malingering of cognitive impairment (1999). Probable is defined as the substantial incentive to fabricate cognitive dysfunction, the presence of evidence strongly suggesting volitional exaggeration or fabrication of cognitive dysfunction, and an absence of credible alternative explanations (p. 182).

Neuropsychological Assessment and Malingering.

Neuropsychology, a specialty within clinical psychology, is the study of the relationship between brain functioning and behavior (Kolb & Whishaw, 1996). A neuropsychologist is trained to obtain quantifiable, normative data about brain-behavior relationships, and therefore possible brain damage and subsequent behavioral consequences (Reitan & Wolfson, 1993). Brain damage or disorder can occur from, but is not limited to: neurotoxicity, tumors, cerebrovascular disease, dementias, and chronic substance abuse. Brain damage can also include acquired brain deficits from incidents such as perinatal damage, various accidents, and on-the-job injuries.

Neuropsychological assessment evolved out of the need to develop tests that, when administered to individuals with identified deficits and compared to individuals with intact functioning, would allow for the inference of what constitutes normal brain-behavior relationships (Reitan & Wolfson, 1993). Additionally, neuropsychological assessment attempts to measure how neurological functions and dysfunctions affect cognitive, emotional and personality functioning. Such assessments can be utilized to determine the presence, severity, and effects of a given injury (Wasyliw & Golden, 1985). Neuropsychological evaluations employ a variety of instruments to assess the degree of deficits ranging from simple (e.g., motor speed) to more complex (e.g., abstraction) cognitive functioning.

In some cases, neuropsychological assessment has been a more sensitive detector of brain damage and deficits than neuroimaging (Barth, Gideon, Sciara, Hulsey & Anchor, 1986). This is particularly evident with closed head injuries, in which the skull remains intact. Closed head injury damage is capable of resulting in moderate to severe cognitive and personality impairment, which may go undetected by tests such as x-rays, MRIs (Binder, 1990), and CT scans. Such damage can be more severe and complex than that inflicted by open head injuries, as the impact of the closed head trauma passes through the intact skull and is absorbed directly by the brain, resulting in shearing, tearing and abrasions (Wasyliw et al, 1985) at coup and contrecoup sites, as well as contusions (bruises) and diffuse axonal injury (Binder, 1990). Coup injuries occur at the site of impact (Lezak, 1983; Kolb & Whishaw, 1996), while contrecoup injury occurs at opposite the site of impact, due to reverberation, or movement of the brain within the skull. As Lezak (1983) states, Athe force of the blow may literally bounce the brain off the opposite side of its bony container, bruising brain tissue...@ (p. 167). Axonal injury is the twisting or shearing of nerve fibers throughout the brain, occurring most commonly in the frontal and temporal lobes (Kolb & Whishaw, 1996). Damage to the fiber tracts can also occur to the corpus callosum and anterior commissure, which interferes with communication between the hemispheres, leading to further deficits. Most impairment results from damage to the frontal and temporal lobes. Frontotemporal damage can result in an inability to inhibit socially unacceptable behavior by loss of control over one=s emotions and behaviors. This loss may be observed through behaviors such as extreme agitation, incoherence, crying spells, uncooperativeness and/or aggressive impulsivity (Lezak, 1983; Kolb & Whishaw, 1996). Likewise, memory functioning can be profoundly affected by closed head injury, and complaints regarding memory dysfunctions are commonly received by clinical neuropsychologists (Brandt, 1988; Bernard & Fowler, 1990). As discussed by Kolb and Whishaw (1996), memory deficits can persist upwards of 1 to 2 years following the incident, and appear to recover slower than intellectual functioning.

Because of the frequent difficulty ascribing etiologies to cognitive deficits, and the inability in most cases to provide physical proof of brain damage through the use of neuroimaging, the issue of malingering becomes very complex and critical. It was postulated in a study by Binder and Willis (1991) that a majority of patients with minor head injuries may be motivated to feign cognitive deficits to obtain financial compensation. Mittenberg, D=Attilio, Gage, and Bass (1990) suggested that the most commonly feigned head trauma may be persistent post-concussion syndrome (PCS). PCS remains one of the most controversial and problematic consequences following closed-head injury, according to Reitan and Wolfson, due to its subjective self-report nature, as well as the difficulty in defining and validating its symptomatology (1986). A study conducted by Mittenberg et al. (1990) identified symptoms which have a high probability of being malingered: AHeadache, depression, anxiety, irritability, difficulty concentrating, fatigue, dizziness, double vision, and light sensitivity@. The authors concluded that malingering must be considered in the absence of convincing evidence of cerebral dysfunction, a presumption which was supported by Binder (1992b), who suggested that malingering may be more common in minor head trauma patients than other, more severe, neuropsychologically-impaired groups. Binder=s contention is that more severe traumatic brain injury patients have little to no difficulty proving deficits and disability (1992a), and therefore little to no need to malinger.

A second, more fundamental problem in the assessment of malingered and factitious neuropsychological impairment is the marked variations and apparent contradictions in the clinical presentation of actual brain-injured patients (Pankratz, 1988). As identified by Reitan et al., headaches and dizziness, or disorders of equilibrium, are respectively the first and second most common symptoms of PCS (1986). In addition to headaches and dizziness, additional neurological and cognitive symptomatology becomes significantly more difficult to identify in routine examinations. A further confound in the diagnosis of malingering is the diversity of psychiatric symptomatology found with head trauma. For example, depressed patients may have scores similar to malingers due to their decreased attention during the assessment process (Pankratz, 1988). In those cases in which a physician is unable to find any medical or physical explanation for the patient=s symptoms, there may be the tendency to label the problem as psychiatric (Ziskin, 1984). Reitan and Wolfson (1985b) reason that, Athe absence of definite neurological findings...has led many neurologists and neurological surgeons to conclude that the patient=s complaints are due to emotional reactions to the head trauma...@ (p. 216).

Furthermore, cognitive deficits can complicate the diagnosis of psychiatric presentations. Criteria for several psychiatric disorders listed in the DSM-IV (APA; 1994), such as schizophrenia, manic and hypomanic episodes, major depressive episodes, dysthymia, posttraumatic stress disorder, generalized anxiety disorder, psychogenic amnesia, and a number of adjustment disorders, include cognitive problems in their diagnostic criteria. Diametrically, some brain-injured individuals can perform within normal limits on any given number of neuropsychological tests examining cognitive functioning, but due to changes in personality or psychological presentation, particularly those related to frontal or temporal lobe damage, are unable to maintain employment or close interpersonal relationships (Larrabee, 1990). Causal factors for deficits can be confounded as well, as neuropsychological testing can indicate deficits in cognitive functioning due to, but not limited to, anxiety, chronic pain, side effects of medication, fatigue, as well as poor motivation (Youngjohn, 1991), which can also give rise to inaccurate testing results. According to Pankratz (1988) "accurate assessment is dependent on patient cooperation because neuropsychological techniques mostly measure behaviors that can be consciously modified (p. 169)."

Studies Utilizing Neuropsychological Assessment Batteries

In this section, studies will be presented that examine malingering using either neuropsychological assessment and/or a neurologically disordered population. Most studies utilize a simulation design as defined by Resnick (1992) as the fabrication of a disorder by an individual who does not possess it. Additionally, most researchers required the simulating group to feign believable disorders, therefore meeting either the criteria for Roger=s (1988a) definition of moderate malingering or Resnick=s (1992) definitions of pure and partial malingering. Initially, studies utilizing neuropsychological batteries for malingering detection will be reviewed. Secondly, studies utilizing one psychological instrument for malingering detection will be reviewed, followed by a review of studies utilizing more than one psychological instrument in the experimental design. Finally, studies utilizing the Rey 15-Item Memory Test (RMT; Andre Rey, 1964) will be reviewed. Although there are a number of instruments designed specifically to screen for psychological malingering, they have not been validated against actual neuropsychological deficits. In contrast, the RMT is a neuropsychological instrument that was specifically designed to screen for malingered amnesia.

First to be examined are two neuropsychological batteries used widely in clinical practice: The Luria-Nebraska Neuropsychological Test Battery (LNNB; Golden, 1981) and The Halstead-Reitan Neuropsychological Test Battery (HRNB; Reitan & Wolfson, 1993). Both are bereft of validity or malingering scales. Additionally, very few studies utilizing these batteries have been conducted that attempt to discriminate malingerers from honest responders.

Luria-Nebraska Neuropsychological Battery (LNNB)

The LNNB (Golden, 1981; Lezak, 1983) consists of 269 items, each of which epitomizes a distinctive type of skill, such as rhythm differentiation. The items differ from each other in terms of stimulus utilized, response required, complexity, and challenge, with scores produced for accuracy, speed, and quality or number of responses. There are eleven primary summary scores, consisting of the following: (a) motor functions, (b) rhythm, (c) tactile functions, (d) visual functions, (e) receptive speech, (f) expressive speech, (g) writing, (h) reading, (i) arithmetic, (j) memory, and (k) intellectual processes. Furthermore, three summary scores are computed: (a) a right-hemisphere score, (b) a left-hemisphere score, and (c) a pathognomonic score. The test takes approximately 2 to 2-1/2 hours to administer. In their interpretive manual, Moses, Golden, Wilkening, McKay, and Ariel (1983) addressed the issue of malingering on the LNNB. They presented three techniques that can assist a clinician in the determination of malingering. Retesting using the LNNB was seen as the best way to identify a malingerer. Moses et al. (1983) noted that malingerers who are retested seem incapable of producing the same profile, with subsequent testing resulting in either much more positive or much worse scores. A second technique utilizes interpretation of deficit severity. The authors point out that in a number of instances, level of deficits are simply incompatible with observed behavior outside of testing, including interaction ability with other medical staff. Finally, production of confusing test results or profiles should alert the examiner to the possibility of malingering. The authors do caution that the last technique may be difficult to utilize due to its level of sophistication and need for an experienced examiner.

Only one published study using the LNNB was found using a PSYCHLit search. In a 1986 study, Mensch and Woods tested psychologically naive participants on the LNNB, motivating them to appear brain damaged by promising a small financial reward if they were successful. The participants (*N* = 32) included 16 High Average IQ and 16 Average IQ volunteers recruited through postings in various employment settings. Each participant was administered the LNNB twice, with a minimum of 14 days separating the first and second administration. Instructions were counterbalanced and randomly assigned via sealed envelopes to permit for blind testing. Half the participants were instructed to feign during the first administration and the remaining half instructed to feign during the second. For all participants, results obtained under malingering instructions produced higher scores than results under honest instructions, which produced normal scores. Although malingered scores resulted in scale elevations, the feigned performance was not consistent within the battery, showing deficits on specific tests and not general ones (e.g. sensory/motor deficits), results which support Binder=s (1992a) contention that malingerers tend to fake selectively. Participants produced scores of 60T or higher on nine of the 30 factors, with the highest mean elevation on Simple Tactile Sensation (*M* = 88T), followed by Number Reading (*M* = 70T) and Simple Verbal Arithmetic (*M* = 68T). Additionally, though the participants were able to produce elevations higher on one hemisphere, extreme differences required to indicate a lateralized disorder were produced in only 10 cases.

Halstead-Reitan Neuropsychological Battery (HRNB)

The HRNB (Reitan & Wolfson, 1993) was designed to address a conceptual model of brain functioning. It consists of six categories that measure both simple and complex cognitive tasks: A(a) input measures; (b) tests of attention, concentration, and memory; (c) tests of verbal abilities; (d) measures of spatial, sequential, and manipulatory abilities; (e) tests of abstraction, reasoning, logical analysis, and concept formation; and (f) output measures@ (Reitan & Wolfson, 1985a). The input category is the registration of the presented stimulation to the brain, while the output category assesses the effector organs through motor functioning, such as Finger Oscillation (finger tapping). Studies investigating the diagnostic ability of the HRNB will be examined initially, followed by studies looking individually at the Finger Oscillation Test and the Booklet Category Test.

To determine the severity of feigned deficits, Prigatano, Samson, Lamb, and Bortz (1997) examined the hypothesis that suspected malingerers perform at lower levels on the Digit Memory Test (DMT) than patients with brain dysfunction. In phase one, individuals referred for neuropsychological evaluation (*N* = 21) were administered the DMT, as well as the BNI Screen for Higher Cerebral Functioning (Barrow Neurological Institute; named for the institute at which the test was developed). The BNI Screen was developed to examine Aabilities across multiple cognitive functions@, and is utilized Aas an indicator of overall level of cognitive functioning@ (p. 612). The 21 patients were assigned to one of three groups dependent upon their diagnosis: 10 patients had moderate to severe traumatic brain injury (TBI), five patients had history of temporal lobe damage (TL), and six patients were suspected of malingering (MAL) various disorders (i.e. post-concussion syndrome, possible cerebral anoxia). The determination of suspected malingering was based on the following two criteria as outlined by Lamb and Prigatano (as cited in Prigatano, Samson, Lamb & Bortz, 1997), which were implemented for the six patients in question: A(1) the patient=s subjective complaints of neuropsychological impairment far exceed what would normally be expected given their medical history, and (2) the level and pattern of neuropsychological test performance are highly unusual given the patient=s history@ (p. 611).

In phase two, 25 subjects with varied brain etiologies (i.e. dementia of the Alzheimer=s type, tumors, TBI) were compared to the initial 21 participants. Additionally, five malingerers from a previous study by Prigatano and Amin (1993) were added to the malingering group in this study for analysis purposes. Results indicate that across the board, with the exception of the participants with Dementia of the Alzheimer=s Type (DAT), the malingerers= performance (*M* scores = 91.8, 85.5, and 82 across three sets of the DMT) was consistently poorer than all other participants who performed >95% on all sets, suggesting that the DMT has clinical usefulness in the detection of suspected malingering. However, caution must be exercised as patients with a dementing process may fail the DMT, obtaining scores at or below the malingering participants.

The severity of deficits presented by malingerers was also examined by Heaton, et al., (1978) who found results similar to the Prigatano, et al., (1997) study. Heaton and his colleagues (1978) administered the complete adult version of the HRNB, as well as the WAIS and MMPI. The results from 16 volunteers (*M* age = 26.7) who were instructed to malinger were compared with the results from 16 nonlitigating head-trauma patients (*M* age = 24.4). The head-trauma participants had documented histories of head injuries, and at least 12 hours of unconsciousness. None were involved in civil or criminal litigation, nor were they applying for disability support. Findings indicate that the malingering group did as poorly as the actual head-injured group in overall test performance, as indicated on the WAIS verbal, performance, and full scale IQs, and on the neuropsychological summary measures on the HRNB (impairment ratings and index). The malingering participants also performed at significantly lower levels on the Finger Oscillation Test. Interestingly, the two groups differed in the pattern of strengths and deficits. According to Heaton et al., the malingerers did poorly on motor and sensory tests, but relatively well on several cognitive tests most sensitive to brain damage. Specifically, the malingerers performed poorly on Speech-Sounds, Finger Oscillation, finger agnosia, sensory suppressions, hand dynamometer, and WAIS digit span tests, and displayed significantly more personality disturbance on the MMPI clinical scales and higher elevations on the F scale. The head-injured participants performed significantly worse on Category Test, Trails B, and Tactile Performance Test total time, memory and location. Using two stepwise discriminant function analyses, the researchers were able to create discriminant function formulas; one using the neuropsychological variables; the other using the MMPI variables. Neuropsychological variables consisted of Category Test errors, Trailmaking Test, Parts A and B total time and errors, Tactile Performance Test total time, memory and location, Speech Sounds Perception errors, Seashore Rhythm Test correct answers, Finger Oscillation Test total taps for 20 seconds, Tactile Form Recognition total time, Finger Agnosia errors, Finger Tip Writing errors, Suppressions errors, Crosses rating and Aphasia errors. MMPI variables consisted of T scores from the three validity scales and ten clinical scales. From all assessment variables (*N* = 37), the analysis identified 30 variables that correctly classified 100% of the participants from both groups. The researchers then cross-validated the discriminant functions using archival data files from 84 previously tested patients. Forty-two of the 84 files were involved in court cases or had given evidence of exaggeration during testing; of these 42 files, 27 (64.3%) were classified as malingerers. Of the remaining 42 files which were not in litigation nor had evidence of exaggeration, only 11 (26.2%) were classified as malingerers.

An additional arm of the Heaton et al. study was to present the post-test protocols to a selected group of ten neuropsychologists, who would act as judges and provide independent Ablind@ ratings as to the authenticity of each protocol, with the instructions that some of the protocols were malingered. Additionally, they were asked to rate their confidence in their decisions, based on a 4-point Likert scale: very sure, sure, fairly sure, unsure. The neuropsychologists accurately classified 50% to 68.8% of the participants. Heaton et al found that, Asensitivity, or true positive rate for real head injuries ranged from 43.8% to 81.3%@, while Aspecificity, or true negative rate for malingerers ranged from 25% to 81.3%@ (p. 895). Therefore the accuracy of the neuropsychologists ranged from chance to approximately 20% greater than chance. The correlations of confidence ratings for accurate versus inaccurate judgments ranged from -.13 to .46, with the more experienced neuropsychologists having higher accuracy in their predictions of correct classification. The correlation between years of experience and confidence ratings was .24.

Likewise, Goebel (1983) achieved success in determining malingering utilizing neuropsychological judgment when administering the HRNB to brain-impaired patients (*N* = 52) and normal college and community volunteers (*N* = 202). The nonimpaired participants were assigned randomly to either a control group, or one of four faking groups. Each faking group differed in the form of brain damage malingered (nonspecific, right, left and diffuse). All protocols were then intermingled, with identifying information removed. Each protocol consisted of the neuropsychological summary sheet, the Aphasia Screening and Sensory-Perceptual examination sheets, and the sheet which contained the participant=s results from the Aphasia Screening. Using clinical judgment, Goebel, the sole rater, sorted the protocols into two groups (brain injured or malingered). He achieved a hit rate of 94.4% in detection of malingerers, with a false positive rate of 5.6% and a false negative rate of 5.7%, with 80% of the patient group correctly classified. Additionally, using the Impairment Index value of .4 as a ceiling limit for normal, a hit rate of 86.2% was achieved. However, the Index also achieved a high false positive rate of 36.7%, and a false negative rate of 3.7%.

Trueblood and Binder (1997) also sought to investigate neuropsychologists= accuracy in detecting malingering based on test data, and obtained positive, but less successful results than Goebel (1983). Protocols of four malingerers, and two severely head-injured individuals were mailed to 440 psychologists randomly chosen from neuropsychology membership directories.

The malingering cases had previously been identified as having produced below chance results on forced-choice testing (FCT). Only two of the malingering cases mailed included the results from FCT. For all cases assessment instruments included the HRNB, as well as several other tests (i.e., WAIS-R, WRAT-R, WMS, Rey AVLT, WAIS-R, Grooved Pegboard, Face-hand Test.) The data mailed to the respondents included brief histories, medical information, hand dominance, educational level, and whether or not the protocol was in litigation. Each psychologist received two cases, either malingering cases or head-injured cases. In a cover letter indicating the detection of malingering was the primary focus, psychologists were asked to choose one of the cases to review. Out of 100 completed questionnaires, 86 were included for analysis in this study. Across the four malingered protocols, error rates (diagnosis of cerebral dysfunction due to head injury) for the psychologists ranged from 0 to 25%, however, the majority of the psychologists were able to distinguish the head injured protocols from the malingered protocols. Although experience level was not discussed, confidence level of accuracy was determined via ANOVA analysis. Findings indicate that confidence level was significantly affected by the inclusion of FCT data (M = 3.72, σ = 0.93), less so when FCT data was excluded (M = 2.85, σ = 0.93).

Using the concept of neuropsychological judgment, Faust and his colleagues performed two studies examining the HRNB using a younger population; the first involved three adolescent malingerers, aged 15-17 (Faust, Hart, Guilmette & Arkes, 1988), and the second involved three pediatric malingerers, aged 9-12 (Faust, Hart & Guilmette, 1988). In the first study, the ability of select neuropsychologists to detect malingering was examined by providing them test results from the HRNB and a fabricated history of mild to moderate head injury. No respondent detected malingering. In the second study, the ability of select neuropsychologists to detect malingering was examined by providing them a summary of the test data, the answer sheets and drawings from the Aphasia Screen. Again, no respondent detected malingering. The mean level of clinical experience for the psychologists in the first study was 8.4 years, while in the second study it was 8.57 years. According to the authors, in neither study did success in detection Avary in relation to training and experience@ (p. 510.) This contention is consistent with Heaton et al., (1978) results which indicate that clinical experience is unrelated to malingering detection. Faust, Hart, and Guilmette (1988) postulate that being unable to select preferred instruments, make behavioral and testing observations, and conduct an interview, the psychologists may not have had favorable conditions for malingering detection.

In the final two studies examined, cross validated step-wise discriminant function analyses were conducted on HRNB results. Mittenberg, Rotholc, Russell and Heilbronner (1996) administered the HRNB to two groups: (a) a validation group consisting of 40 normal volunteers and 40 non-litigating head injured patients, and (b) a cross validation group consisting of an additional 40 normal volunteers and 40 nonlitigating head injured patients. All groups were matched on age, gender, and Impairment Index. MANOVA analyses indicate that significant differences between groups were found in both the validation [$F(13, 66) = 5.28$, $p<.001$] and cross validation sample [$F(13, 66) = 13.40$, $p<.001$]. A step-wise discriminant function analysis formula was developed to aid in malingering determination. The formula utilized raw scores from the Category Test, Tactual Performance Test, Seashore Rhythm Test, Speech-Sounds Perception Test, Trail Making Test, Finger Oscillation Test, and Perceptual Disorders Examination, as well as scoring ceilings for the Tactile Performance Test (10 minutes per trial), Trailmaking Test, Part A (180") and Trailmaking Test, Part B (300"). Subtest scores were entered into the formula for both the validation and cross validation samples, accurately classifying 88.75% of the validation and 83.75% of the cross validation participants. True positives for both samples were 92.5% and

87.5%, while true negatives were 85% and 80% for the validation and cross validation samples, respectively. The authors report that based on analysis which was not identified, classification accuracy was not related to age, gender, Impairment Index, educational level, length of time since injury, or length of unconsciousness. Overall, the patient sample performed worse on the Tactile Performance Test total time and Trailmaking Test, Part B time, while the malingering sample performed more poorly on the Category Test, the Speech-Sounds Perception Test, the Seashore Rhythm Test, sensory suppressions, and finger agnosia.

Utilizing the analysis formula from the Mittenberg et al. (1996) study, McKinzey and Russell (1997) obtained a high number of false positives (27%) in 796 participants being evaluated with the HRNB for treatment purposes. No participant was involved in litigation and therefore had no motive for malingering. All questionable profiles indicative of faking or malingering were excluded from the study. Of the 120 head trauma cases, 22.5% were incorrectly classified as malingerers, while 62% of the patients with moderate to severe Impairment Rating were classified as false positives. Obviously, this formula needs significant improvement, however the authors suggest it may be a helpful adjunct to the screening process.

Studies Utilizing Multiple Neuropsychological Assessment Instruments.

Research by Johnson and Lesniak-Karpiak (1997) sought to investigate the effect of warning regarding detection of malingering on memory, motor, and verbal tasks. Participants were general psychology undergraduates (*N* = 87), with no differences in age or education between groups. The three groups consisted of simulators with warning, simulators without warning, and control subjects. All participants were administered the Grooved Pegboard, the Wechsler Memory Scale-Revised and the Verbal subtests from the WAIS-R. The WAIS-R was administered to control for variance in IQ. Following a two hour testing session, a post assessment questionnaire was administered to determine understanding and compliance with instructions. Results indicate that a warning prior to performance was effective in reducing malingering on memory and motor tests, with simulators without warning performing significantly worse than simulators with warning. However, significant differences were not found on Attention/Concentration, Visual Memory, and dominant/nondominant hand motor performance. The authors suggest that provision of a warning prior to testing as an adjunct to existing detection techniques may reduce the probability of malingering behavior through discouragement. One method by which warning regarding malingering can be provided is through the informed consent for assessment.

The utilization of classification rates was examined by Davis, King, Klebe, Bajszar, Bloodworth & Wallick (1997) who utilized statistical analyses to determine malingering, obtaining good hit rates with a population of 120 undergraduate volunteers (simulating group = 60; control group = 60). The two groups did not differ on age, education, IQ or gender. All participants completed two priming tests. The first priming test consisted of a list of 15 words presented on a computer, which the participant was to rate according to emotional context toward the word (i.e. 1 - dislike very much, 5 = like very much). Participants were to then complete 20 three letter word stems presented one at a time, saying aloud the first word that popped into their head, as well as typing in the remaining letters, with a maximum of ten seconds allowed for responding. Ten of the word stems were from the rating portion of the test, while the remaining 10 were from a frequency matched list. The response latency of the participants was then recorded. In the second priming task, (the juice test) which was administered after a 30 minute break following the first task, the procedure was identical; however, the words were different. Stem words that had only one completion (Jui = juice) were utilized. Again, latencies were measured. The experimenters administering both tasks were blind to the instructions. Priming scores for each participant were calculated for each task, as well as response latency scores. Based on discriminant function analysis, 80% of the non-malingering control group and 73.3% of the simulated malingering group were correctly classified, with an overall hit rate of 76.7%. The simulating group had lower priming scores than the control group [F (1,118) = 42.4, p<.001], as well as larger mean latency scores than the control group *[*F (1,118) = 55.5, p<.001].

The Rey 15-Item Memory Test (RMT) Studies

Several studies have been completed which examine the accuracy of the RMT in determining malingering in simulators, patients, and suspected malingerers. The majority of the findings indicate that malingerers do tend to produce lower cut-off scores than controls and most populations of actual neurological patients. However, the results are not consistent, as the majority of the cut-off scores for the malingering sample fall between 7 and 11. Additionally, researchers differ on the cut-off score to be utilized in determining malingering, with different studies utilizing different cut-off scores. Due to the number of studies examining the RMT, they are described in an individual fashion below based upon their results.

In examining the RMT, Bernard (1990) discovered that it Awas not found to be related to the determination of malingering.@ Bernard assessed malingered memory deficits utilizing the Wechsler Memory Scale-Revised (WMS-R), Complex Figure Test (CFT), Auditory Verbal Learning Test (AVLT), and RMT. All instruments were administered to three groups of undergraduate students in psychology. The control group (N = 28) responded honestly, while the two other groups were instructed to malinger; malingering with a financial incentive (N = 30), and malingering without a financial incentive (N = 28). Results indicate there was not a significant difference on the RMT total number between groups, and only minimal significance on number of sets between groups. Although the malingering groups responded in like fashion, their results were significantly poorer than the control group. Specifically, there was not a significant difference between the groups on the RMT, however, the simulators could be distinguished from the controls on recognition versus recall tasks (WMS-R Figural Memory, CRT Recall, and AVLT Recognition). Bernard suggests that the RMT may have been more successful if given at the beginning, rather than the end, of his testing battery.

Unlike his first study, in a follow-up 1990 study by Bernard and his colleague Fowler, the RMT was found to exclude 88.8 % of the brain-injured participants and 100% of the controls. It was administered to 18 brain-damaged, and 16 volunteer controls (comparison group) who were paid $5.00 for their participation. The brain-damaged group was administered the RMT, the WAIS-R, the Complex Figure Test (CFT), and the Auditory Verbal Learning Test (AVLT). The control group was administered the RMT, the Shipley Institute of Living Scale, the CFT, and the AVLT. The two groups were matched via frequency distribution on age, education and sex. Results indicated no significance in the number of sets recalled for the two groups. However, the total number of items recalled was significant. This study is especially relevant in that it validated the cutoff score on the RMT as 9 items for suspicions of malingering, as suggested by Lezak (1983). The authors further suggest a more conservative cutting score of 8, which would have excluded 100% of the participants in this study.

Bernard again sought to examine the efficacy of the RMT with a simulation population, but obtained mixed results. In a 1993 study with his colleagues, Houston and Natoli, he examined the RMT along with Hebb=s Recurring Digits (HRD), the AVLT, the WMS-R, and the CFT. All instruments were administered to two groups (controls and simulated malingering) of 57 randomly assigned undergraduate students. The two groups were not significant for age or gender. The tests were given in the following order to all participants: RMT, CFT Copy, WMS-R, CFT Recall, AVLT, and HRD. The results showed significantly lower mean scores for the simulated malingering group in all cases. The RMT discriminated between simulated malingerers and controls, but exceeded the recommended cut-off of 9. For this study, the mean for total number recalled was 10, with a mean of 2.2 sets recalled. Bernard et al. (1993) recommend further studies utilizing the RMT with malingerers.

Schretlen, Brandt, Kraft, and Van Corp (1991) found that the RMT has limitations as they addressed the issue of feigned amnesia by administering the RMT to 76 participants feigning (69 participants faking amnesia or insanity and 7 suspected malingerers), 148 patients with amnesia or another neuropsychiatric disorder (10 amnesiacs, 40 patients with mixed neuropsychiatric disorders, 55 patients with traumatic brain injury, 9 patients with mixed dementia, 34 severely mentally ill patients), and 80 control participants who responded honestly. Using the less than 9 items as the cutoff score, the RMT had a predictive value of 21.6% for both the fakers and patients. Many simulators and suspected malingerers recalled at least nine items or three categories, which is above Lezak=s (1983) suggested cutting scores. However, many of the patients (particularly those with dementia, amnesia, and severe mental illness) fell below the cutoff score.

Scoring on items only, Greiffenstein, Baker, and Gola (1994) achieved an acceptable hit rate as well with the RMT. They administered the RMT, Rey Word Recognition List (WRL), Reliable Digit Span, Portland Digit Recognition Test and Rey AVLT to examine three groups. The first group consisted of 33 traumatic brain injury (TBI) individuals. The second group consisted of 30 persistent postconcussive syndrome (PPCS) individuals. The last group consisted of 43 probable malingerers (PM; who were PPCS patients which met criteria for malingering. Interrater reliability between two of the researchers in classifying the PPCS individuals into those with and those without malingering symptoms was 90%. The RMT correctly classified 71% for the

TBI/PM comparison, and obtained group differences for total items recalled for the TBI (*M* = 12.8), PPCS (*M* = 13.2) and PM (*M* = 9.7) participants.

In their second study, this time examining feigned amnesia, limitations were again found as Greiffenstein, Baker, and Gola (1996) administered the RMT and the Rey WRL to 60 patients with traumatic brain-injury (TBI), and 90 litigating post-concussion patients who were probable malingerers (PM). The WRL administration consisted of having the client listen to a list of 15 words, then provided a typed list of 30 words: the original 15 words, and 15 foils. The client was instructed to circle the original 15 words. Clients with reading difficulties were administered the test verbally. The authors examined eight different methods of scoring the RMT and WRL, and found that correct WRL responses produced better hit rates, and appear to be more sensitive than the RMT for malingered amnesia.

The opposite results were found in a 1994 study by Simon examining the utility of the RMT in malingering assessment. Simon (1994) administered it to incarcerated male defendants (*N* = 14) who had been suspected of malingering, and a randomly selected control group (*N* = 14) which consisted of patients from the same facility who had been acquitted by reason of Amental disease or defect@ (p. 915). The results were scored by two Ph.D.-level psychology interns. Reliability coefficients between the two interns were .98 for total number of RMT items correctly recalled, and 1.00 for total number of rows correctly recalled. The results were highly significant, with control participants recalling more RMT items (*M* = 10.07) than the malingerers (*M* = 4.57). Recall of sets, or rows, was also significant, with the control group recalling 2.07 mean sets, and the malingerers recalling .71 mean sets. Using Lezak=s (1983) cutting score of nine, 85.7% of the malingerers would have been correctly identified, while only 14.3% of the controls would be false-positives.

A reduction in the cut-off score (7) by Lee et al., met with mixed results in a 1992 study utilizing the RMT to collect data on 100 temporal lobe epilepsy (TLE) inpatients without litigation, and 56 neurologically-impaired outpatients with litigation (*N* = 16) and without litigation (*N* = 40). Disorders in the outpatient sample included tumors, ruptured aneurysms, closed and penetrating head injuries, and hydrocephalus. Results indicate that TLE inpatients performed comparatively with outpatients without litigation, yet significantly better than outpatients with litigation. However, even with these results, the majority of outpatients with litigation scored within the normal range (62.5%), and above the suggested cutting score of seven.

Low hit rates were obtained in a final study examining the feigning of cognitive symptoms by worker=s compensation claimants by Boone, Savodnik, Ghaffarian, Lee, Freeman and Berman (1995). They administered the RMT, the Dot Counting Test, and the Millon Clinical Multiaxil Inventory (MCMI) to 154 litigating individuals who claimed psychiatric injury for the co-occurrence of (1) feigning of cognitive and psychological symptoms and (2) feigning of cognitive symptoms associated with personality dysfunction. The injuries were secondary to work related stressors, as the claimants presented for a psychiatric evaluation per the request of worker=s compensation insurance providers. On the cognitive malingering tasks, 12% of the participants obtained scores of less than 9 in the RMT (*N* = 7), failed the Dot Counting test (n = 16), or both malingering instruments (*N* = 4). Statistical analyses indicated that the subgroup that faked on both the cognitive tests and MCMI differed significantly from the other two groups, with a higher incidence of exaggeration of psychiatric symptoms (*N* = 15%), rather than faking of psychological symptoms (*N* = 8%). Interestingly, those individuals who failed the malingering tasks endorsed more traits that reflect personality disorder, as well as psychotic features, in relation to the remainder of the sample. However, the cognitive faking rate is low, and suggests that only about one-fourth of participants in a worker=s compensation population who feign psychological symptoms also feign cognitive symptoms.

Studies Utilizing Individual Neuropsychological Assessment Instruments

On most instruments, naive simulators are not able to successfully duplicate the results of actual patients. Typically, simulators either greatly exaggerate symptoms, or provide significantly different cutting scores than honest, neuropsychologically injured responders. Bruhn and Reed (1975) found that normal college students were not successful in feigning organicity on the Bender-Gestalt test when reviewed by both an expert and non-expert clinician, and a college senior. The simulators provided results which differed significantly from organic subjects, producing patterns which were easily identified. Furthermore, organic subjects typically are the only patient population which experiences intersection difficulties on card 6, as well as rotations (Lezak, 1983).

In contrast, a 1961 study by Benton and Spreen obtained a lower mean score for 70 simulators than for 48 brain-damaged patients on the Benton Visual-Retention Test (BVRT). This test is utilized in neurological or psychiatric settings to assist in the detection, and extent, of brain damage. Sixty-six percent of the simulators, and only 27% of the brain-damaged participants obtained scores of 0-2 correct. The mean score for brain-damaged participants was 3.73, while that of the simulators was 2.16. The simulating group made more errors of distortion, while the patients made more omissions, perseverations, and alterations in size.

Likewise, malingerers scored below control and brain-damaged participants in a study designed by Iverson, Franzen, and McCracken (1991). They hypothesized that malingering individuals would perform significantly worse than both normal and memory impaired individuals. 20 normal controls and 20 memory malingerers randomly selected from a group of college undergraduates, and 20 memory impaired patients were examined. All participants were administered a 21-item word list containing object nouns. The word list was read to each subject, followed by a recognition task with 21 additional words as distractors. The simulating malingerers were provided a case scenario in which they were married with two children and a drinking problem, committed DWI, and killed another driver. They were to feign amnesia to decrease their penalty or be found not guilty. On both tasks, the control participants outperformed both the malingerers and actual memory impaired patients. Based on the results, a cutting score of 9 for the forced choice recognition task was determined to minimize false positives, which resulted in the correct classification of 100% for control and memory impaired individuals, 65% for malingerers, with an overall hit rate of 88%. The recall portion of the task was less effective in correctly identifying members of each of the three groups, mislabeling a number of memory impaired patients.

Finger Oscillation Studies

Studies examining the performance of malingerers on the Finger Oscillation test of the HRNB have obtained similar results. Heaton et al. (1978) showed that malingerers could be distinguished from nonlitigating head trauma patients using the HRNB. Specifically, the malingerers provided poorer performance on the Finger Oscillation Test, as well as other measures on the HRNB. Similarly, Mittenberg et al. (1996), found that malingerers performed significantly worse on the Finger Oscillation Test (M = 63) than nonmalingering head trauma patients (M = 75.64). Mittenberg et al.=s finger tapping score was calculated as the sum of the average performance of the right and left hand trials (1996). These results are consistent with Binder et al.=s (1991) previous findings.

Prigatano et al. (1993), obtained similar findings, administering Finger Oscillation to examine the efficacy of the Digit Memory Test (DMT), a forced-choice test for malingering detection. Participants included patients with neuropsychological disorders (*N* = 32), suspected malingerers (*N* = 6), and controls (*N* = 10). No significant difference was found between either the patients (*M* right = 44; *M* left = 42), or suspected malingerers (*M* right = 49; *M* left = 39) on finger tapping. However, it should be noted that there was a greater than 10% left-right difference (*M* > 20%) for the suspected malingering group, which merits further testing.

Booklet Category Test Studies

The few malingering detection studies that have been conducted with the Booklet Category Test present with conflicting results.

Heaton et al. (1978) used results from the HRNB, WAIS, and MMPI to compare 16 volunteer malingerers and 16 nonlitigating head-trauma patients. Although no significant differences were found between the groups, patterns of strengths and deficits produced differed between malingerers and head-injured patients, with nonlitigating head trauma patients performing significantly worse on the Category Test than malingerers.

In contrast, Mittenberg et al.=s 1996 study found that head trauma patients obtained fewer total errors (*M* = 62.69) than the malingering group (*M* = 83.19).

Wisconsin Card Sorting Test Studies

The ability of the Wisconsin Card Sorting Test (WCST) to correctly classify participants as controls, malingerers, psychiatric and brain-impaired populations was nonconclusive in a study by Knight, Webster, Goetsch, Malloy, and Greve (1986). They administered the WCST to four groups to assess its= ability to discriminate malingerers. The groups consisted of 34 psychiatric inpatients, 62 brain-impaired individuals, 31 controls, and 58 subjects instructed to malinger. Results indicated significant differences between the malingerers and remaining three groups on the number of categories obtained, total number of errors, total correct responses, and number of nonperseverative errors. However, no analyses were performed to determine the classification accuracy based on the scores obtained.

Symptom Validity Testing Studies

In a survey oriented experimental design, psychological ignorance of brain damage was tested by Mittenberg et al. (1990), who provided 197 community participants with a list of 50 affective, somatic and memory complaints and asked the subjects to determine which symptoms were likely to be malingered following head injury. The participants were also asked to imagine that they were involved in litigation following a motor vehicle accident, and were to produce symptoms in order to increase compensation. Repeated measures ANOVA results indicated a significant post injury symptom increase ($F(1,116)=205.98$, $p<.0011$), with post-hoc Bonferroni comparisons indicating that PCS symptoms were more often malingered than memory deficits (p=s<.001). Additionally, actual memory symptoms which accompany cerebral trauma (i.e. anteriograde verbal and nonverbal memory deficits) were malingered about half the time; however, remote and procedural memory symptoms which are not present in concussion cases were also malingered, such as inability to remember addresses, birthdays and telephone numbers.

In a second study assessing malingering of PCS, Cradock and Gfeller (1996) assessed the ability of sixty undergraduate students to feign PCS symptoms, obtaining results consistent with Mittenberg et al.=s 1990 study. Cradock and Gfeller randomly assigned students to one of three groups: an honest control group, a naive simulating group, or a sophisticated simulating group. All students were ultimately compared to a group of 20 patients with PCS symptomatology, who were three months post injury due to motor vehicle accidents (1996). Both groups of simulating students were instructed to malinger believable symptoms associated with traumatic brain injury, with the sophisticated group receiving additional information about cognitive, emotional and physical sequelae. Mean total scores and summary scores for symptom frequency, intensity and duration from the Postconcussion Symptom Checklist (PCSC) were calculated. Results indicated that the control group received significantly lower scores than the two malingering and one patient group. Both simulating groups received scores comparable to the actual patient group, with no significant between group differences noted. These findings indicate successful symptom presentation by the simulating groups.

In studies which assess the symptomatology of a disorder, such as the feigned post-concussion syndrome discussed in Mittenberg et al. (1990) and Cradock et al. (1996), utilization of symptom validity testing has produced significant results, and is promising in its use for malingering detection. A major drawback to the use of SVT in clinical practice is the development and administration of a personalized test targeted for specific symptoms which are suspect of being malingered by a client or patient.

In a study examining the simulation of neuropsychological deficits utilizing SVT, Frederick and Foster (1991) used SVT with a nonverbal test of cognition (TONI) as a means to detect malingering. Unlike standard SVT, they reordered the 50 items on both parts A and B, to incorporate increasing item difficulty. Participants included 84 simulated malingers (SM) of college age, 86 normal controls (NC) also of college age, and 14 male forensic psychiatric patients without pending adjudication made up the cognitively impaired (CI) control subjects. The SM students were asked to feign brain damage. Seventy-three percent of the students from the SM group were still identified as malingerers, even after obtaining accurate information on detection avoidance.

Even more significant than examining the efficacy of SVT in a study, is the ability of SVT to detect individuals in real life who feign or exaggerate symptomatology. The following are actual cases, in which SVT was utilized to detect the suspected malingerer. Pankratz (1983) assessed three patients complaining of memory deficits. In all cases, following the presentation of symptom validity testing, each patient appeared to improve, showing mild to marked decreases in memory deficits. The SVT involved using two lights, one red and one white, which were illuminated in random order. One light was illuminated for two seconds, following which the patient was to perform the Digit Modalities Test for 15 seconds and then asked to remember which light was illuminated. In Case 1, the patient accurately recalled all 25 trials. In Case 2, the patient obtained correct responses on all 20 trials. In Case 3, the patient missed five out of the first 20; however, thereafter correctly recalled the remaining trials (number not noted).

Again utilizing the single case design, Binder and Pankratz (1987) also obtained less than chance results, using the SVT technique in the assessment of a 53-year-old female, who was being evaluated for Social Security Disability. Previous claims based on physical and psychological deficits had been denied by medical and mental health professionals. Additionally, neurological evaluations, including a CT scan were normal. The neuropsychological assessment administered by the researchers indicated severe faking. Therefore, the authors devised a symptom validity test to assess her memory complaints. A black pen or yellow pencil were presented to her and removed, after which she counted from 1 to 20. She was then asked to recall which object she had seen. She made 63 errors on 100 trials, a definitely less than chance score.

Below chance results were obtained by Hiscock and Hiscock (1989), who outlined a case history of a 45-year-old male who claimed inability to perform his employment responsibilities due to impaired memory from a head injury incurred seven years prior. The subject, a restaurant owner, was struck by a motor vehicle, sustained a right frontal lobe fracture, and was rendered unconscious for three days. He subsequently claimed anosmia, facial numbness, hearing loss, headaches, and seizures, along with the inability to remember short lists or his customers' faces. The examination performed by the first author, was at the request of the workers= compensation board. After formal testing was consistent with the hypothesis that the patient was malingering, the authors specifically designed a forced-choice memory test. The patient=s performance (29%) fell significantly below chance level, and supported the hypothesis of malingering.

Tests Utilized Specifically for Malingering

Aside from the positive findings utilizing chance performance, the accuracy of classification rates in the determination of malingering has also been examined. Binks, Gouvier and Waters (1997), administered the Dot Counting Test (DCT) to 93 participants. The DCT was conceived by Rey and has been reported by Lezak (1983) to be a useful test for malingering detection. Two packets of six cards present sets of ungrouped and grouped dots, respectively. The participant is to count the dots as quickly as possible, with patterns of counting times for grouped dots expected to be shorter than ungrouped dots. Additionally, the total number of incorrect responses was recorded. The 93 participants were undergraduates who were assigned to one of three groups with equal participants ($N = 31$): Naive simulator, sophisticated simulator, and control. A fourth group, a clinical sample of neuropsychological patients, consisted of 26 participants who were not involved in any litigation. The sophisticated simulators were provided strategies or techniques on how to simulate without detection. Results indicates that a high number of errors for the sophisticated group were misclassified into the naive group, as were most of the errors for the neuropsychological patient group classified into the control group. Therefore, the four groups were collapsed into simulating and non-simulating groups. Based upon these two groups, the overall correct classification rate was 85%, with the correct classification of 89% for non-simulators, and 81% for simulators.

Bickart et al, (1991) examined the Malingered Memory Deficit Test (MMDT), a newly devised instrument for the detection of malingering of memory complaints. The MMDT was administered to 114 male inmates of the Kentucky State Reformatory. All participants were volunteers who gave written informed consent, were compensated for their performance with packages of cigarettes, and were all serving sentences on felony convictions. The test was administered in a group format of 19 to 26 participants. The MMDT was a 50-trial task which required the participant to Aview a sequence of consonants, consisting of a pair of letters in random sequence (e.g., RTRRT)@ (Bickart et al., 1991, p. 7). After a delay of five seconds, the participant is show a second sequence of consonants identical to the first, with the exception of three or five missing consonants (e.g., R_RRT), for which they were to supply the missing consonants. To further enhance the study, task interpolation was introduced by asking the participant to count the number of elements in random configuration. Therefore Athere were four versions of the MMDT: Difficult/Interpolation, Difficult/No interpolation, Easy/Interpolation, and Easy/No interpolation@ (Bickart et al., 1991, p. 7). Using valid positive cutoff scores of 20 or fewer correct responses, results indicated that about 80% of the malingerers were successful in escaping detection by the MMDT. Additionally, more than a quarter of the participants produced results which were classified as chance responding (scores between 21 and 29).

Summary

Studies of single assessment instruments or symptom surveys to accurately detect malingered neurological and/or neuropsychological deficits show mixed results. Bruhn and Reed (1975) found that non brain-damaged college students were not successful in feigning organicity on the Bender-Gestalt test when reviewed by an expert and non-expert clinician and a college senior. Conflicting results were found in a 1961 study by Benton and Spreen, in which the mean score for 70 simulators (M = 2.16) indicated more impairment than the mean score for 48 established brain-damaged patients (M = 2.16) on the Benton Visual-Retention Test (BVRT). Malingerers again scored below control and brain-damaged participants in a study designed by Iverson, Franzen, and McCracken (1991). In a study examining the efficacy of the WCST to correctly classify participants as controls, malingerers, psychiatric and brain-impaired populations, Knight, Webster, Goetsch, Malloy, and Greve (1986), received mixed results. However, no analyses were performed to determine the classification accuracy based on the scores obtained.

In a survey oriented experimental design, psychological ignorance of brain damage was tested by Mittenberg et al. (1990), who provided 197 community participants with a list of 50 affective, somatic and memory complaints and to identify which symptoms were likely to be malingered following head injury. Results indicated a significant post injury symptom increase ($F(1,116)=205.98$, $p<.0011$), with post-hoc testing indicating that post-concussion symptoms were more often malingered than memory deficits (p=s<.001). In a second study assessing neuropsychological malingering simulation of post-concussion syndrome, Cradock and Gfeller (1996) obtained results consistent with Mittenberg et al.=s 1990 study.

Frederick and Foster (1991) used SVT with a nonverbal test of cognition (TONI), with 84 simulated malingerers, 86 normal controls, and 14 male forensic psychiatric patients without pending adjudication. Seventy-three percent of the simulating malingerers were still identified as malingerers. In a single case design, SVT has proven to be highly effective in the detection of suspected malingering. Pankratz (1983) utilized SVT to assess three patients complaining of memory deficits. In Case 1, the patient accurately recalled all 25 trials. In Case 2, the patient obtained correct responses on all 20 trials. In Case 3, the patient missed five out of the first 20;

however, thereafter correctly recalled the remaining trials. Binder and Pankratz (1987) also obtained less than chance results in the assessment of a 53-year-old female, who was being evaluated for Social Security Disability. She made 63 errors on 100 trials, a definitely less than chance score. Finally, Hiscock and Hiscock (1989) assessed a 45-year-old male who claimed an impaired memory from a head injury seven years prior. The patient=s performance (29%) fell significantly below chance level, and supported the hypothesis of malingering.

The accuracy of classification rates in the determination of malingering has mixed results as well. Binks, Gouvier and Waters (1997), assessed 93 undergraduates, who were ultimately assigned to a simulating or non-simulating group. The overall correct classification rate was 85%, with the correct classification of 89% for non-simulators, and 81% for simulators. Finally, Bickart et al, (1991) examined the Malingered Memory Deficit Test (MMDT), utilizing an incarcerated population. Using valid positive cutoff scores of 20 or fewer correct responses, results indicated that about 80% of the malingerers were successful in escaping detection by the MMDT.

In conclusion, while some of the instruments were more successful than others in detecting malingering, the general findings indicate that neuropsychological malingering is a very complex issue, with a large number of symptoms which may or may not be correlated or interrated. Even those instruments which have been created specifically for malingering detection obtained either mixed or less than favorable results. These studies advocate the need for the continued development of assessment instruments which are more effective in the accurate detection of neuropsychological malingering. In the following section, recommendations and techniques from psychologists and mental health professionals which may assist in the detection of malingering will be summarized. However, these guidelines are more observations, and although they may assist clinicians in determining a malingering diagnosis, there is still a need for standardized testing instruments which utilize classification and hit rates, as well as cut-off scores in the detection of malingering.

Guidelines for the Detection of Malingering

Currently, there is limited research that directly examines clinicians' capacity to detect feigned testing results obtained with neuropsychological examinations and batteries. The previous research has shown that unless the clinician possesses significant experience with neuropsychological malingering, the use of clinical judgment alone as the predictor of malingering provides poor detection rates (Heaton et al., 1978). Youngjohn (1991) states that clinicians= interpretation of response patterns and inconsistencies are neither conclusive, nor very accurate, and further expounds that it is possible that subjective inconsistencies may occur with patients with actual deficits. Heaton et al. (1978), Faust, Hart, and Guilmette (1988), and Faust, Hart, Guilmette, and Arkes (1988) all illustrated in their studies that blind clinical judgment and cutting scores also are not accurate methods for determining dissimulation. Of significant importance is the fact that not only can deficits on effortful tasks be malingered, but involuntary neurological responses such as the patellar reflex can be brought under voluntary control to fool examiners (Stam, Speelman, and van Crevel, 1989, secondary quote in Binder, 1992a).

The detection of malingering of such deficits continues to be a difficult and controversial issue, regardless of the type of malingering being assessed. Current controversy arises from the lack of a universally accepted classificatory model of malingering, as well as definition of malingering, as discussed previously. Faust and Guilmette (1990) postulate that contributions to a neuropsychologists= inability to accurately detect malingering may include the lack of definitive criteria for identifying malingerers, as well as the lack of reliable base rates of malingering within clinical populations which impedes detection as clinicians do not know how often to expect deception. Therefore, there is also controversy in identifying an individual a malingerer when he/she is not, while for those patients which do feign deficits, few will admit to malingering in a clinical situation (Mittenberg et al., 1996). Conversely, Youngjohn postulates that reasons for the difficulty arise from issues such as poor detection instruments, to additional reasons, aside from intentional malingering, why an particular individual may perform poorly (1991). According to Gorman (1984), identifying malingering on an examination Arequires proof that no disorder exists that (1) accounts for the sign or symptom, and (2) such a disorder is under the individual=s volition@ (p. 72). Inaccurate clinical interpretations regarding severity of the injury, patient motivation, or genuine psychological symptomatology can occur (Binder, 1990). In cases of suspected malingering, the DSM-IV (APA, 1994) provides guidelines to assist in detection. The guidelines include (a) referral by an attorney, (b) significant discrepancy between the deficits and clinical findings, (c) poor cooperation and compliance with both assessment and treatment, and (d) the diagnosis of Antisocial Personality Disorder (APD). However, these criteria have been criticized by clinicians (Rogers, 1990s, 1990b, 1997) for expecting them to determine internal, as opposed to external, motivation and for unjustly overemphasizing the occurrence of malingering with APD. Youngjohn summarizes the general viewpoint regarding malingering detection by stating that Awhen the motivational context of the assessment is such that the patient stands to gain from poor performance, the possibility of malingering must be considered@ (1991).

To address the above issues, a number of researchers have submitted that certain tests be utilized in the detection of neuropsychological malingering (Wasyliw et al., 1985). Lezak (1983) suggests that a number of tests appear much more difficult than they are; therefore, malingerers should perform more poorly than the majority of brain-injured patients, a strategy which utilizes floor effect, and will be discussed later in this research. According to Bash and Alpert (1980) malingerers present with an over endorsement of symptomatology, as well as unique patterns of responding, theories expounded upon by Rogers (1984), and Rogers et al. (1993). In their 1980 study, Bash et al. identified answers on the subtests of the WAIS-R that would be considered approximate, and were endorsed much more by suspected malingerers than actual psychiatric inpatients. This detection strategy, titled Amagnitude of error@ by Rogers et al. (1993) will be examined in this research. According to Rogers (1988b), the extent to which any response is inaccurate may be far more revealing than a simple correct/incorrect answer in determining level or magnitude of dissimulation.

There are general questions which, when answered, may assist in the diagnosis of malingering. Pankratz (1988) devised the following as useful criteria for determining when to assess for malingering. First, is the client involved in any litigation or criminal proceedings? Second, can secondary gains be received by the client from having a deficit? Third, does the client have a history of malingering or factitious disorder? According to Lezak, patient history is as important as testing performance (1983). Even with these questions, a determination should not be ascertained from an interview alone. An analysis of consistency and inconsistency within testing performance can be used for the detection of factors other than brain injury deficits (Brandt, 1988; Pankratz, 1988). This strategy can be accomplished in four ways, (a) evaluation of consistency between tests can be performed, (b) test performance can be compared to established deficits for neuropsychological disorders, (c) test performance can be evaluated in relation to the severity of the brain injury, and (d) consistency of the deficits and behavior of the individual can be evaluated. Specifically, the evaluation of both within and between test consistencies can be evaluated, as well as examining whether the performance is consistent with established patterns for known disorders. Evaluation of the performance relative to the severity of the injury can also be analyzed. Finally, the individual=s behavior and its= consistency with testing performance can be evaluated. Pankratz (1988) subsequently points out that the best information can be gleaned from Adeficit testing@, rather than generalized brain damage testing, a theory supported by Binder (1992a), who states that malingers tend to fake selectively, rather than mimicking global, severe mental impairment. Pankratz states that the malingering individual has difficulty reproducing deficits correlated with the disorder he/she is seeking to fake, therefore utilizing individual tests versus neuropsychological batteries has advantages geared toward symptom testing (1988).

To further clarify diagnostic clues in the detection of malingering, Pankratz (1988) devised the following model which provides criteria which suggest malingering. The individual being assessed need only endorse one of the following criteria:

> (a) near misses to simple questions, (b) gross discrepancies from expected norms, (c) inconsistency between present diagnosis and neuropsychological findings, (d) inconsistency between reported and observed symptoms, (e) resistance, avoidance, or bizarre responses on standard tests, (f) marked discrepancies on test findings that measure similar cognitive ability, and (g) failure on any specific measure of neuropsychological faking@.

Franzen et al. (1990) suggest a screening procedure to detect malingering as a beginning part of the assessment process. A screening procedure can also assist in determining the need for additional assessment instruments, or testing procedures (such as symptom validity testing), which may more specifically address the issue of malingering. This could be particularly helpful in any setting in which clients or patients may potentially benefit from appearing psychologically or cognitive impaired.

A final consideration in malingering detection focuses upon addressing patient presentation during an assessment procedure (Pankratz, 1979; 1988). He states that malingering patients usually present with significant anxiety that can occur at any time during the testing session. This anxiety typically presents as resistance and can be found through excessive talking which delays the testing, or through arguments about the test design, or inability to complete the task. A very common form of resistance is to violate the test rules, such as turning away from the assessment instrument, or responding with the same answer throughout the assessment process. A final response of an anxious or resistant individual is to explain the success of their correct answers.

Six Specific Strategies Employed to Detect Malingering

Specific strategies employed in the detection of neuropsychological malingering include: (a) floor effect; (b) performance curve; (c) magnitude of error; (d) symptom validity testing, (e) atypical presentation; and (f) psychological sequelae. These strategies will be briefly described below.

Floor effect. This strategy describes the Afloor@ or lower limit of a set of scores (Vogt, 1993). The term floor effect implies that the malingerer is failing at tasks on which even grossly impaired individuals are likely to succeed. In his dissertation research, Powell (1991) found that simulating malingerers had more difficulty with easy items, which he attributed to the malingerers= finding it more difficult to construct inaccurate responses when the correct answers were more obvious. According to Larrabee (1990), the tests which best delineate floor effect are multiple trial, attentional, and verbal learning tasks. The RMT is one of the most commonly used measures for floor effect.

Performance curve. A second strategy is to utilize instruments present questions that hierarchally increase in difficulty. The strategy postulates that as questions on neuropsychological tests become increasing difficult, the performance of a individual with brain injury typically decreases; hence, correct responses decrease with increased item difficulty. When a line is plotted to display performance, an individual responding honestly will have a downward slope, which represents a decline in performance. This strategy assumes that malingerers will not take increasing item difficulty into account; therefore, failing easy items and passing difficult ones. Their performance curve is hypothesized to look more like a straight line, with minimal downward slope. This strategy has been successfully applied to memory tasks and intellectual measures.

Magnitude of error. A third strategy, which was conceived by Richard Rogers, Ph.D., from the University of North Texas, focuses on evaluating the quantitative features of incorrect responding. It summarizes that the magnitude of incorrect responding (the degree of wrong) can be an indicator of the level of malingering or motivation to malinger. Unsophisticated simulators are more likely to provide answers which are significantly unrelated to the correct response, sometimes even bizarre in content, whereas individuals with disorders such as dementia are more likely to provide responses which closely resemble the correct answer, rather than being significantly different. It is hoped that this strategy might yield different patterns of incorrect responses among simulators, from approximate answers to those which are grossly incorrect.

Symptom validity testing. There appear to be no clear indicators for the determination of neuropsychological malingering during the interview or testing session, with the possible exception of symptom validity testing (SVT). SVT is a term coined by Muriel Lezak (Pankratz, 1988) to describe a technique which assesses the validity of symptoms entailing perception and short-term memory complaints (Lezak, 1983). Past research has utilized this technique in the assessment of Ablindness, color blindness, tunnel vision, blurry vision, deafness, anesthesias, and memory loss@ (Bickart, Meyer, & Connell, 1991; Binder, 1990; Binder & Pankratz, 1987; Hiscock & Hiscock, 1989; Pankratz, 1979, 1983, 1988). In using this technique, each test must be individualized exactly for the patient=s complaint, as well as to anticipate the answers of the disinclined patient. Therefore, the testing is presented as an assessment of whatever disability the patient claims to have.

Measures incorporating symptom validity testing rely on the application of a two-alternative, forced-choice technique (Pankratz, 1983). According to Pankratz (1988), the examiner must identify a replicable stimulus which corresponds with the disability. This stimulus is then presented for approximately 100 trials, utilizing the two-alternative, forced-choice method (e.g., yes-no answers). By chance alone, approximately 50% of the subject's answers will be correct, which is the expected result with a valid complaint. The task is presented as a straightforward assessment of the claimed disability (Lezak, 1983).

Research on malingering has provided statistically significant results using forced choice testing. According to researchers (Binder, 1990; Binder and Pankratz, 1987; Hiscock and Hiscock, 1989, Lezak, 1983), the malingering individual typically performs significantly worse than chance, as they try too hard. Binder and Pankratz (1987) report that a score significantly below chance level is more indicative of an attempt to fake. Also, according to Pankratz, in most clinical situations, malingering individuals too frequently Aguess@ incorrectly, as they perceive a 50% hit rate as too successful (1983). Furthermore, an important aspect of symptom validity testing is the feedback to the patient, which is used to reinforce the Agood@ performance, even when the patient is not performing well. Additionally, those patients who fear detection may admit they can perceive or remember the stimuli (Pankratz, 1979).

In utilizing SVT in the detection of malingering, Larrabee (1990) notes two inherent weaknesses: (a) it is time-consuming, needing a considerable number of trials, and (b) it is unable to provide additional information to the neuropsychological exam. An additional weakness is that a majority of participants are able to discern the pattern and nature of the technique, and therefore change their answers. Finally, it must be considered that an individual failing SVT may actually be impaired, and other testing data must be interpreted in such a light. In response to the patient=s ability to discern the pattern, Binder (1990) developed the Portland Digit Recognition Test (PDRT), which appears to be significantly more difficult to malinger on than other versions of SVT.

Atypical presentation. Inconsistent or atypical performances on tests of similar abilities or additional presentations of the same test are believed by some researchers to be indicative of malingering. Commonly, neuropsychologists look for patterns across instruments which measure the same constructs or abilities in determining the presence or severity of brain damage. If inconsistencies occur, the performance of the participant falls into question. Likewise, consistent performance over several testing sessions enhances the validity of the testing protocol. However, researchers such as Pankratz (1988) suggest that inconsistent or atypical presentation is not uncommon in brain-injured patients. As Rogers et al. (1993) point out, Athe virtual absence of empirical data on atypical presentations as an indicator of malingering militates against the use of this strategy in the determination of feigning@ (p. 262), yet, further stipulate that such inconsistencies may warrant further investigation.

Psychological sequelae. This strategy examines the co-occurrence of cognitive deficits and psychological symptoms. A number of researchers have found that simulators of brain-injuries and/or physical complaints report a greater number of psychological symptoms than actual brain-injured individuals (Heaton et al., 1978; Lees-Haley, 1990; Mittenberg et al., 1990; Pankratz, 1988). However, the lack of empirical research investigating the use of psychological symptom endorsement in the determination of malingering makes this strategy troublesome. Further research is warranted.

The Purpose of this Study

The purpose of the current research is to obtain guidelines which will allow for better identification of malingering, utilizing a select group of neuropsychological measures. Accordingly, this study will examine the ability of six neuropsychological tests to assist in the detection of a feigned disorder. Several factors led to the selection of these particular assessment instruments. First, given their common use in neuropsychological evaluations, any malingering standards could be utilized by a large proportion of clinicians. Second, they are well-known instruments and their familiarity to clinicians will allow for each of use in malingering assessments. In addition, normal and brain-damaged assessment profiles have been established on these instruments, which will assist clinicians in the detection of irregular results or profiles. Third, as these instruments are commonly used in neuropsychology, they have a high likelihood of being involved in litigation as part of a battery to determine extent of an injury. Therefore, standardization for their use in the identification of malingering will allow for validated testing results to support the professional opinions of clinicians.

As discussed by Pankratz (1988), a single instrument can evaluate only a small number of skills; therefore, we will utilize six instruments that evaluate various functions mediated by the brain. The following functions will be analyzed: As discussed by Pankratz (1988), a single instrument can evaluate only a small number of skills; therefore, we will utilize six instruments that evaluate various functions mediated by the brain. The following functions will be analyzed: Expressive vocabulary via the Wechsler Adult Intelligence Scale-Revised® Neuropsychological Instrument Vocabulary subtest (WAIS-R® NI; Psychological Corporation, San Antonio, TX, www.psychcorp.com); verbal fluency via the Controlled Oral Word Association Test (COWA;

Bechtoldt, Benton, & Fogel, 1962), motor speed via the HRNB Finger Oscillation Test (Halstead & Wolfson, 1993), fund of general knowledge via the WAIS-R NI Information subtest (WAIS-R® NI); abstract thinking, problem-solving, and executive functioning via the Wisconsin Card Sorting Test™ (WCST™; Psychological Assessment Resources, Lutz, FL, www.parinc.com), the HRNB Booklet Category Test (Halstead & Wolfson, 1993); and the WAIS-R NI Similarities subtest (WAIS-R® NI); memory via the Rey-15 Item Memory Test (RMT; Andre Rey, 1964) and the WAIS-R NI Information subtest (WAIS-R® NI); and general intelligence via the WAIS-R NI Vocabulary subtest (WAIS-R® NI).

The current study will utilize Resnick=s conceptualization of pure (falsification of all symptoms) and partial (exaggeration of deficits) malingering (1992). These definitions parallel Rogers= (1988a) conceptualization of moderate malingering, which involves exaggeration or fabrication of deficits. The current research utilizes a simulation design, and the simulators will be asked to address two criteria: (1) the concept of pure malingering by creating head-injury symptoms where none exist, and (2) the concepts of partial and moderate malingering through the exaggeration or fabrication of deficits that are neither too mild nor too severe to create a believable disorder. Simulation is defined by Resnick (1992) as positive malingering, in which an individual feigns symptoms or a disorder that he or she does not possess; a definition which is typically utilized for research purposes.

It is hoped that the design of the present research will permit a number of issues to be addressed. First, and foremost, the ability of the current study to obtain successfully malingered profiles will help ascertain whether one can malinger on neuropsychological assessment instruments. Specifically, we will examine whether specific neuropsychological deficits and/or symptoms can be feigned, in contrast to general, diffuse deficits. The effectiveness of the malingering profiles will also be examined via comparison to those of actual brain-injured patients. Second, we will examine whether scores and/or profiles of brain-injured participants with litigation are significantly different from brain-injured participants without litigation. Third, cutting scores will be developed based on the consistent within-group performance of the simulators, as well as significant differences between the simulating and nonsimulating groups. Finally, based on group frequency means, false negative and false positive error rates, as well as positive predictive value (PPV) and negative predictive value (NPV) percentages will be developed to assist in the detection of malingering. Therefore, the implications of the findings of the present study will be important for technicians and professionals in neuropsychological assessment, particularly those involved in forensic issues in neuropsychology.

Lastly, Rogers recommends that (a) participants asked to malinger should be provided an incentive for success, (b) malingering instructions must be explicit and underscore the believability of the malingered behavior, and (c) compliance of the participants must be periodically verified (1988b). Rogers et al. (1993) indicate that without efforts to measure incentives and their effects upon the participants, as well as the participants motivation to follow instructions and consistently attempt to malinger believable deficits, the results from any research has limited generalizability to real-world malingerers. As suggested by Berry, Lamb, Wetter, Baer and Widiger (1994), particular types of coaching may have a significant influence on the success of malingers. One method for verifying a participant=s motivation to follow instructions and feign a deficit is through the use of debriefing. The recommendations put forth by Rogers (1988b) will be heeded, as malingering instructions, incentives for participation, and post-testing debriefing will be in the current study.

Estimates of Diagnostic Accuracy

The current research is intent on determining cut-off scores, false negative and false positive error rates, and PPV and NPV on six neuropsychological measures to assist in the identification of malingering. Towards that end, the concepts of sensitivity, specificity, PPV, and NPV as they apply to the validity of the assessment instruments will be examined to better clarify the accuracy of such instruments in their classification of malingering identification. These terms were defined by both Rogers (1995), and by Johnson, Hamer, Nora, and Tan (1997). Sensitivity and specificity are methods of reporting the efficacy of an assessment instrument, and are dramatically influenced by the base rate. Explicitly, the term sensitivity refers to the ability of each assessment instrument to correctly identify the number of participants malingering, as identified by a positive score indicative of a disorder or deficit. A high sensitivity percentage indicates an instrument that can identify a significant number of true positives. Generally speaking, how accurate is the measure in identifying individuals who have a particular disorder? Conversely, the term specificity refers to the ability of each assessment instrument to correctly identify the number of participants who are not malingering; therefore, are responding honestly, as identified by scores which fall within the normal limits. High specificity percentages typically indicate an instrument will yield positive results when a person has the particular disorder or deficit the instrument predicts.

Unlike identification via an assessment instrument, PPV and NPV address the identification of the individual, and provide information as to what the results of the assessment instrument indicate. PPV indicates estimates regarding the probability an individual is a malingerer, or more generally, the percentage of participants with a positive score who have the disorder or deficit. NPV indicates estimates regarding the probability an individual is a non-malingerer or honest responder, or more generally, the percentage of participants with a negative score who have no disorder or deficit. A high NPV implies that if an instrument says someone is not a malingerer, chances are pretty significant that he is not.

Research Hypotheses

The first hypothesis is that participants who malinger tend to produce poorer scores on the RMT, COWA, Finger Oscillation, and WAIS-R NI than populations of head-injured and control participants alike, yet produce scores similar to the forensic participants. Likewise, it is hypothesized that participants who malinger tend to produce better scores on the WCST and the BCT than populations of head-injured and control participants, yet produce scores similar to the forensic participants. This hypothesis will utilize concept of floor effect in the statistical analysis. Determination of sensitivity, specificity, PPV, NPV, and cutting scores will also be functions of this hypothesis.

The second hypothesis is that the magnitude of incorrect responses (the degree of wrong) can be an indicator of the level of malingering or motivation to malinger, with the malingering group producing a greater degree of wrongness than the remaining four groups. To test this hypothesis, an average magnitude of error per incorrect answer score on the WAIS-R NI Vocabulary, Information, and Similarities subtests, the BCT, and WCST was analyzed.

The third hypothesis is that the rate of decay (accuracy) in responding on the COWA, Finger Oscillation, and WAIS-R NI tests for the malingering group would be smaller from the rate of decay for the head injured and control groups, yet be similar to the forensic groups. This hypothesis implied that participants who malinger tend to have a lower rate of decay on assessment instruments, which could be plotted using the concept of performance curve. Typically, the performance curve of a malingerer is indicative of a straight line, unlike honest responders who tend to have a decline in performance, therefore a downward curve. It has been reasoned that malingerers not only miss earlier, easier items, they also have difficulty decreasing their performance with increased item difficulty.

The fourth hypothesis will examine the efficacy of Bolter=s 14-item Performance Validity Index (VI) from his 1992 study. The revised BCT answer sheet illustrating the VI will be examined to assess its viability in successfully detecting neuropsychological malingering. It is hypothesized that the simulating group would obtain more errors overall on the BCT, and therefore, would be more likely to obtain a higher number of VI errors than the control, head-injured with litigation, head-injured without litigation, and forensic groups.

CHAPTER II

METHOD

Participants

The current research consisted of five groups of 25-26 participants each. The sample sizes were estimated based upon a power analysis using Cohen=s rule of thumb for a medium effect size of .5 (Cohen, 1992; Kirk, 1990). The acceptable power level was .80, with an alpha of .05 (Faul & Erdfelder, 1992). Utilizing this effect size estimate, statistically interpretable differences were expected. Standardized group means and variances were unavailable for some assessment instruments; therefore, an a priori power analysis for some instruments could not be computed prior to statistical analysis. However, retrospective power analyses have been computed for each hypothesis to aid in the interpretation of the statistical analyses.

The control group consisted of participants who were asked to respond honestly. The simulating group consisted of participants who were instructed to malinger. Participants for the control and simulating groups were randomly assigned via a sealed envelope which the participants chose from a larger envelope. The sealed envelope provided directions and an informed consent form pertinent to group membership. The participants were instructed to read and sign the directions and informed consent, then place them back in the envelope, re-seal it, and inform the examiner when they were ready to proceed. Instructions for both groups prohibited participants from informing the examiner which group he or she was in. Specific instructions for the control group asked participants to respond in an honest nature. Conversely, instructions for the simulating group asked participants to feign a head injury from an automobile accident (see Appendix C).

The third group consisted of patients with mild to moderate head injuries who were involved in litigation. The fourth group consisted of patients with mild to moderate head injuries who were not involved in litigation. Groups three and four were obtained from Neuropsychology Associates of Dallas, a private neuropsychology practice that specializes in neurocognitive assessment. These two groups were included in the current study to provide profiles that illustrate a neurologically dysfunctional sample of the population. They were split into two groups, litigating and nonlitigating, to allow for examination of the influence of possible or pending financial gain on assessment results. The litigating head-injured group was also included to provide a scoring guideline for the malingering group, and to assist in the possible determination of the type and extent of exaggeration or fabrication in a genuine head-injured, possibly financially motivated sample of the population.

The fifth group consisted of adult males on parole for felony offenses, whose parole officers were employed by the Volunteers of American Ohio River Valley, Inc. The inclusion of this group was based mainly on the premise that the diagnosis of malingering in the forensic population is high (Rogers, 1986), and will therefore permit a comparison between perceived malingerers and simulators. Exclusion for participation in the forensic group was recent illicit drug use. To that end, all participants were required to have passed their two most recent urine analyses, with the most recent analysis performed on the morning of the testing.

Participants in the control, simulation and felony groups were recruited both by board postings at the University of North Texas and through solicitation by the primary researcher. Participants who gave informed consent (Appendixes A, B and D) in accordance with both the University of North Texas and Volunteers of America Ohio River Valley, Inc. ethic/review board guidelines were briefed on the purpose of the study and were evaluated during a single session, typically 1-1/2 to 2 hours in duration. The head-injury group participants were evaluated during a standard neuropsychological battery, therefore consent was not obtained. For all groups, researchers attempted to control for the variables of age, gender, ethnicity, and estimated IQ. An IQ estimate was calculated from the Vocabulary subtest on the Wechsler Adult Intelligence Scale-Revised® Neuropsychological Instrument Vocabulary subtest (WAIS-R® NI; Psychological Corporation, San Antonio, TX, www.psychcorp.com), which was administered during the testing session.

The current research utilized a simulation design in contrast to a “known-groups” design. In a simulation design, Anormal@ participants are given instructions by the examiner to feign a neuropsychological impairment and are compared to both (a) other Anormal@ participants with honest instructions, and (b) one or more comparison groups (e.g. forensic) that were also given honest instructions (Rogers et al., 1993). A major benefit of simulation designs is the use of experimental controls and comparison groups (Rogers et al., 1993). Additionally, simulation designs have the purpose of assisting in the identification of individuals who are suspected of malingering brain deficits. However, this reasoning assumes that the behavior and response styles of malingerers are comparable to those of normal participants feigning head-injury. The realistic comparison between these two groups is unknown. An additional concern is the motivation level and subsequent performance of the simulators. As noted previously, malingerers are often experienced deceivers who have much to lose if they are detected (Binder, 1992a). In contrast, simulators are often college students or members of the facility or community where the researchers live and work, and are used as volunteers who have little or no incentive to malinger. Therefore, malingerers have an investment in the success of their performance which a simulator does not. Further, it is unlikely that simulators have interacted with numerous health care professionals or have obtained a significant amount of knowledge regarding neuropsychological deficits and assessment instruments. On the contrary, the malingerers are likely to have been in the mental health system, filling out checklists and answering questions that describe neuropsychological symptoms.

Conversely, in a “known-groups” design, identified malingerers or individuals who are highly suspected of fabricating or exaggerating disability (Binder et al., 1991) are compared with patients and/or controls. An inherent problem with a “known-groups” design is the lack of uncontaminated and reliable criteria for malingering (Greiffenstein et al., 1996). However, as mentioned by Rogers et al. (1993), the “known-groups” design is directly applicable to genuine malingerers.

We used a simulation design in the current study to allow for control of the specific type of injury and situation that was feigned by the simulation group. A simulation design also circumvents the difficulty of finding and identifying known malingerers. The use of simulators has also been consistent in studies investigating the efficacy of neuropsychological test batteries in determining feigning.

Rogers= (1988b) recommendations to improve simulation research mentioned previously were implemented in the current study. First, the malingering instructions were specific and clear and emphasized the simulation of a Abelievable@ disorder. Second, the ability of the participants to follow instructions was analyzed through a Amanipulation check@ or post-assessment questionnaire. Finally, incentives were provided for the control, simulating, and forensic groups for their participation ($10.00 each). In addition, a financial reward of $50.00 was given to a participant from either the control or simulating group following the completion of data analysis. This participant was randomly chosen from a lottery, in which the code numbers for all control and simulating participants were placed during data collection.

Materials

Each participant was administered all measures in the following order: Rey-15 Item Memory Test (RMT; Andre Rey, 1964); Controlled Oral Word Association Test (COWA; Bechtoldt, Benton, & Fogel, 1962); HRNB Finger Oscillation Test (Halstead & Wolfson, 1993); HRNB Booklet Category Test (Halstead & Wolfson, 1993); Wisconsin Card Sorting Test$^{\text{TM}}$ (WCST$^{\text{TM}}$; Psychological Assessment Resources, Lutz, FL, www.parinc.com); and the Wechsler Adult Intelligence Scale-Revised® Neuropsychological Instrument Vocabulary, Information and Similarities subtests (WAIS-R® NI; Psychological Corporation, San Antonio, TX, www.psychcorp.com).

Additionally, a handedness inventory, the Lateral Dominance Examination from the HRNB (Halstead & Wolfson, 1993), was administered to determine the hand dominance of each participant. Instruments that were found in past studies to be successful in the detection of malingering were not utilized in the current study. The purpose of this exclusion was the need to develop standards on instruments for which they currently do not exist. Rogers= (1990) position that, A...there are no psychological measures which are not fakeable,@ suggests that all standardized psychological and neuropsychological instruments need to studied for the development of malingering indicators. As such a project is beyond the scope of this research, the instruments chosen for inclusion in the current study met the criteria of: (a) being neuropsychologically utilized and appropriate, (b) possessing ease of administration, (c) being easily transportable, and (d) having not been involved in malingering research that obtained successful results. Furthermore, as these instruments are typically administered as part of a neuropsychological evaluation, their use in malingering detection would be cost effective and time efficient, and with the exception of the RMT, would not be readily identifiable as malingering measures. As noted with the RMT, a considerable amount of research has been performed examining its ability to detect malingering, but with inconsistent results.

The Rey 15-Item Memory Test (RMT)

This test was devised by Andre Rey (1964) to validate memory problems, as well as to assist in the detection of minimal effort and/or malingering. The RMT tests short term memory and measures floor effect. It consists of 15 characters and shapes arranged in five rows of three items, printed on a stimulus card. The rows consist of capital letters A, B, C, numbers 1, 2, 3, small letters a, b, c, Roman numerals I, II, III, and geometric shapes of a circle, square and triangle. The stimulus card is exposed for ten seconds. The card is then removed, and the subject is asked to draw what he/she recalls. The task was designed to aid recall, yet is set up to appear difficult by emphasizing that there are 15 separate items to be remembered. In reality, the participant only has to recall five rows of Achunked@ information. Two scores are determined: (a) the total number of items recalled, regardless of order; and (b) the number of correctly ordered rows. Rey (1964) and Lezak (1983) suggested that malingerers would be misled to perform more poorly than impaired patients. Lezak suggested a cutting score of 9 items recalled for the detection of malingering. Research findings for this instrument are controversial; however, with cutting scores ranging from 7 to 11 items recalled. Goldberg and Miller’s 1986 study revealed that mentally retarded participants were unlikely to remember nine items, and commonly made errors of perseverations and reversals - indicators of mental inadequacy.

Reliability and validity of the RMT. Although a literature review on the RMT is included in the introduction section, the results from those studies is briefly summarized below to aid the reader.

Interrater reliability coefficients of .98 for items and 1.00 for number of rows was obtained by Simon (1994), who used a cutting score of 9 items recalled. He correctly classified 85.7% of the simulating participants (*M* items = 4.57), all of whom were incarcerated males suspected of malingering. However, he obtained a 14.3% false positive rate for the control participants (*M* = 10.07), all of whom were patients found not guilty by reason of insanity.

Also using a cutting score of 9, Bernard et al. (1990) correctly classified 88.8% of their brain-injured population, and 100% of their controls. Reducing the cutting score to 8 correctly classified 100% of their population. Shretlan et al. (1991) misclassified 21.6% of both their patient and simulating population using a cutting score of 9 items recalled. Correct classification (71%) of traumatic brain injury and probable malingering participants was achieved by Greiffenstein et al. (1994), who reported mean number of items recalled for both the brain-injury (12.8) and malingering (9.7) groups. Using a cutting score of 7 items recalled, Lee et al. (1992) examined 100 temporal lobe patients without litigation, 40 neurologically-impaired without litigation, and 16 neurologically-impaired with litigation. Sixty-two percent of his neurologically-impaired participants with litigation scored in the normal range, obtaining more than 7 correct items. Finally, Boone et al. (1995) administered the RMT to 154 individuals involved in litigation. Only 12% of his sample scored below 9 items recalled.

The Controlled Oral Word Association Test (COWA)

The COWA (Bechtoldt, Benton, & Fogel; 1962) is a three-minute instrument that measures an individual=s word fluency, or ability to provide a list of words beginning with a specified letter. Two versions of the COWA are available; both of which have been restudied with the intent of updating the normative data (Ruff, Light, Parker & Levin; 1996). Version A consists of stimulus letters C, F and L, while version B consists of stimulus letters P, R, and W. The determination of the stimulus letters was based on the number of words found in dictionaries of the English language. Additionally, for each version, the letters provide an increasing level of difficulty in generating words (i.e., C is less difficult than F, which is less difficult than L).

Reliability and validity of the COWA. Ruff et al., (1996) sought to examine the coefficient alpha and test-retest reliabilities for a sample of 360 normal volunteers, aged 16-70 years old. The educational level of the sample ranged from 7 to 22 years. A coefficient alpha of $R = .83$ was obtained by totaling the number of words generated per letter (three letters total) and summing the totals to obtain the COWA total test score. Test-retest reliability ($R = .74$) was significant at the $p < .001$ level, although the mean total number of words obtained from first ($M = 39.7$) and second ($M = 42.5$) testings were significantly different ($p < .001$ level). However, this finding is suggestive of practice effects. In this study, Ruff et. al. (1996) discovered that the majority of errors included repetitions or perseverations of words.

Finger Oscillation Test

The Finger Oscillation test from the HRNB (ANA finger tapping test), is a measure of motor speed and is sensitive to brain injury in the posterior frontal lobe (Reitan & Wolfson, 1993). The examiner utilizes a manual tapper (electric when testing younger children), and only the index finger of both hands. The hand must be held in a flat position, to invoke only index finger movement. First the dominant hand is tested, with the objective of obtaining five consecutive 10-second trials within a five tap range. If the participant is unable to obtain five trials within five taps, the total of eight trials will be averaged for each hand, after discarding the lowest and highest score. Following the third trial for each hand, a rest period is given irregardless of participant fatigue. Upon completion of the test, a dominant-nondominant hand difference is calculated based on the right and left mean scores. In the majority of participants, the dominant hand is expected to be 10% faster than the nondominant hand. If the results fall within the impaired range, the left-right difference can point to location and severity of brain damage.

Reliability and validity of the Finger Oscillation Test. The HRNB manual (Reitan et al., 1993) does not provide studies of validity for the HRNB subtests. It does, however, provide a minimal description of examination of validity, which will be summarized below.

Reitan et al. (1993) compared a group of 50 participants with documented cerebral damage/dysfunction with a group of 50 controls utilizing the HRNB. The control group consisted of normal individuals (24%) and hospitalized patients (76%) without impaired brain functions. Both groups were matched on race and gender, and significantly similar on chronological age and education. Very striking differences were obtained between the mean scores for the two groups for all variables (tests) within the HRNB. In Halstead=s 1947 normative study (cited in Lezak, 1983), he administered the finger tapping test to 29 normal controls, and obtained mean right hand scores (50 taps/10 seconds) and mean left hand scores (45 taps/10 seconds). It should be noted that there were only eight women in this sample, however.

In a second study by Dodrill (1979), 47 men and 47 women were tested utilizing their preferred hands. Mean scores of 55.87 taps were obtained by the men, and 51.08 taps were obtained by the women, with an approximate 5 tap difference between the groups noted. Finally, Mittenberg (1990), states that malingerers tend to do more poorly on finger tapping than controls.

The Booklet Category Test (BCT)

As a part of the HRNB, the BCT (Reitan & Wolfson, 1993) is considered by its creators to be Aprobably the best single test in the [HRNB] in terms of showing the adverse effects of cerebral damage@ (p. 89). This consideration is not surprising as the BCT was developed specifically to determine brain damage (Reitan et al., 1993). Recent research indicates this test is sensitive to cerebral damage throughout the brain, not just the frontal lobe; however, frontal lobe functions, such as abstraction, reasoning, and analytical thinking can be evaluated with the BCT.

BCT instructions inform the participant that the test is divided into seven subtests, with each subtest having a particular theme or strategy from beginning to end. Booklet One contains subtests one through four, and Booklet Two contains subtests five, six and seven. Booklet one is placed in front of the participant, along with a 2" x 8" panel displaying the numbers one, two, three and four in order. The participant is asked to verbally respond with number 1, 2, 3 or 4 to indicate what number the picture on the presented page reminds them of. The participant is allowed only one answer per page. Based on the examiners= Acorrect@ or Aincorrect@ response to the participant=s answer, the participant is expected to systematically test various themes or procedures to correctly identify the visually-presented strategy. Participants are never told why their answers are incorrect, nor are they provided with the theme for any subtest. At the start of each new subtest, the participant is again provided the instructions, being informed that the theme of the new subtest may be the same or different than the previous subtest(s). According to Mittenberg, head-injured individuals perform more poorly on the BCT than malingerers (1990).

Reliability and validity of the BCT. The HRNB Manual (Reitan et al., 1993) does not provide validity studies for the HRNB subtests. It does, however, describe a study examining validity, summarized below.

Reitan et al., (1993) compared a group 50 participants with documented cerebral damage/dysfunction to a group of 50 controls utilizing the HRNB. The control group consisted of normal individuals (24%) and hospitalized patients (76%) without impaired brain functions. Groups were matched on race and gender, and were similar for chronological age and education. One of the most striking between-group differences for the HRNB tests was observed on the

Category Test. Of the 50 pairs of matched subjects, 47 of the controls performed better than their head-injured match.

To examine the alternate-format reliability of the projector and booklet versions of the Category Test, Holtz, Gearhart, and Watson (1996) administered the two versions in a counterbalanced presentation to 30 brain-injured and 30 controls. All participants received both versions, with a mean interval of 5.8 days between administrations. Correlation coefficients were calculated for both groups examining individual subtest comparisons, as well as total scores. Correlations obtained from the brain-injured sample were significantly larger (total score $r = .88$, median subtest $r = .76$) than those obtained from the control sample (total score $r = .42$, median subtest $r = .55$). Although the authors believed that the data supported the use of the BCT, particularly for a brain-injured population, they cautioned readers about the constraints of their sample, which consisted of 60, primarily middle-aged male veterans. Therefore, further research was recommended to support their results.

In a validation study, the Category Test was examined by Bolter (1992), who attempted to derive an index of performance validity by identifying items infrequently missed by brain damaged (n = 55) and normal participants (n = 50). The number of frequently missed items was computed for both groups, and an item was included in the list of infrequently failed items if it obtained a 95% criterion rate. Results yielded 14 items for the Performance Validity Index (VI), that were subsequently identified on the BCT answer sheet via asterisks. Using the validity index, no control subject earned a score greater than one, and 98% of the brain damaged subjects earned a validity index score of two or less. Bolter suggests that earning a validity index score greater than two is highly atypical for both brain damaged and healthy individuals, and should be regarded as highly suspect (1992).

Johnstone et al. (1997) examined the construct validity of the Category Test using 308 patients with cognitive dysfunction. The major premise of their research was to investigate what exactly the Category Test measures; intelligence, and/or reasoning. The patient population was referred for testing over a four year period for assistance in differential diagnosis of neurologic versus psychiatric dysfunction (i.e., CVA, substance abuse, TBI, seizures). To determine convergent and divergent validity, various instruments were included (WAIS-R subtests, BCT, WMS-R Immediate Logical and Visual Memory, Trails A and B times, and TPT total times for all three trials.) Analyses included both: a) Spearman correlations between the Category Test subtests and all other measures, and b) principal-components factor analysis conducted on 308 participants and 27 variables (25 cognitive test scores, plus age and education). The Category Test was found to be distinct from the WAIS-R, Trails B and TPT and to load on factors distinct from the other tests. Factor analysis resulted in 6 factors, of which 3 identified the Category subtests only: Category Subtests 3, 4, and 7 loaded on Factor 3 (Spatial Positioning Reasoning); Category Subtests 5 and 6 loaded on Factor 5 (Propositional Reasoning); and Category Subtests 1 and 2 loaded on Factor 6 (Symbol Recognition/Counting). These results indicate that the Category Test is not a single measure of abstract reasoning, but instead a measure of distinct reasoning processes.

The Wisconsin Card Sorting Test (WCST)

The WCST (Berg, 1948) was developed to measure abstraction ability in healthy individuals, yet is currently gaining popularity as a neuropsychological instrument. It utilizes four stimulus and 64 response cards that display three different dimensions: a) geometric shapes - cross, circle, square, or triangle, b) colors - red, green, blue, or yellow, and c) numbers - one, two, three, or four. The stimulus cards are completely different from each other: (1) card one - one red triangle; (2) card two - two green stars; (3) card three - three yellow crosses; and (4) card four - four blue circles.

The stimulus cards are placed in front of the participant, and the participant is handed the stack of response cards and instructed to place them, one at a time, beneath the stimulus cards, attempting to match to a dimension. After each placement, the participant is informed only whether the choice is right or wrong. Once the participant has made a predetermined number of correct responses for a particular category (color, form, and number, in that order), the principle is changed to the next category. This procedure is repeated until the participant completes a specified number of categories (typically six).

A recent survey indicates that the WCST is used by a majority (73%) of the responding sample of 500 neuropsychologists (Butler, Retzlaff, & Vanderploeg, 1991). The WCST assesses higher-order cognition and nonverbal frontal lobe functioning, and can provide data regarding magnitude of error and symptom validity. Robinson, Heaton, Lehman, and Stilson (1980) examined the ability of the WCST to detect brain lesions. Their results indicated that the WCST tends to be selectively sensitive to frontal lobe dysfunction. In a 1996 study examining the ability of the WCST to discriminate between patient and control groups, Axelrod, Goldman, Heaton, Curtiss, Thompson, Chelune and Kay (1996) found that an undamaged, normally functioning frontal lobe is needed for accurate performance on the WCST. Brain-damaged individuals tend to perseverate on this test (Heaton, 1981). According to Mittenberg (1990), one technique to detect malingering on the WCST is to review Aother@ responses, that may give an indication as to the existence of frontal lobe damage.

The majority of research on the WCST is with known populations. A great deal of research has been conducted examining the WCST with the following populations: attention deficit disordered, bipolar disordered, depressed, neurologically disordered, obsessive-compulsive disordered, psychotic, schizophrenia disordered, and substance abusers. Those studies which utilize malingering designs have been previously presented in the literature review section of the current research.

Reliability and validity of the WCST. Although a number of early studies had been performed on the WCST, there were variations in test administration and scoring, and study descriptions were ambiguous (Heaton, 1981). Additionally, the early studies suggested that the WCST discriminately measured prefrontal functioning, which has not been borne out in recent studies. Therefore, the earlier studies were not outlined here, but the normative study described in the WCST Manual was discussed with regard to test reliability.

In the normative study, 208 patients with structural cerebral lesions and 150 normal controls were administered the WCST, the Wechsler Adult Intelligence Scale, and the HRNB. The patients had various types of brain damage (i.e. tumors, hemorrhage, trauma, meningitis) and were referrals for neuropsychological evaluations at the neuropsychology laboratory at the University of Colorado Health Sciences Center. Based on the diagnoses, patients were then further divided into one of eleven groups, determined by both lesion type and location. Comparisons were made between the lesion location groups (frontal, frontal plus nonfrontal, nonfrontal, diffuse) and the controls, as well as between the total brain damaged group and controls. Using >18 perseverative responses as a cutting score for brain damage, Heaton correctly classified 74% of the brain damaged and 72% of the control participants, with an overall hit rate of 73.2%. Based on his results, Heaton found that two diagnostic predictions may be attempted utilizing the WCST: Apredictions about presence or absence of brain damage, and, given prior evidence that a cerebral lesion is focal, predictions about presence or absence of frontal lobe involvement@ (p. 34). Results also indicated that the perseverative response score obtained greater diagnostic accuracy than the other WCST scores, such as learning to learn or failure to maintain set. It is cautioned that clinicians utilizing the >18 cutting score should be aware that (a) results suggest that a number of focal nonfrontal patients will be misclassified as normal, and (b) a number of normal individuals over the age of 59 will likely be misclassified as well.

Axelrod et al. (1996) examined the discriminability of the WCST=s indices by using the revised 1993 standardization sample developed by Heaton, Chelune, Talley, Kay, and Curtiss, in comparison with its four neurological patient populations (*N* = 343) and normal control group (*N* = 356). The patient populations were diagnosed as having Astructural brain lesions@ (p. 339), and were divided into (a) focal frontal lobe lesions (*n* = 59), (b) frontal and nonfrontal focal lesions (*n* = 53), (c) nonfrontal lobe lesions (*n* = 54), and (d) diffuse lesions (*n* = 177). The lesions were documented through neuroimaging and/or neurosurgery. The WCST was divided into its six dependent measures: errors, perseverative responses, perseverative errors, nonperseverative errors, percent conceptual level, and categories. One-way ANOVAs indicated that the patient groups performed significantly worse than the controls on all measures. Post-hoc Scheffe analyses did not indicate differences between the patient groups; however, patients with frontal lobe, frontal lobe plus nonfrontal, and diffuse lesions made significantly more nonperseverative errors than controls. The accuracy of the WCST measures in classifying group membership was then examined via discriminant function analysis (DFA). The patients were collapsed into one group and compared with controls. Utilizing a stepwise DFA, overall classification of group members resulted in 71% accuracy, with 64% sensitivity and 78% specificity. Negative predictive value was found to be 69%, while positive predictive value was 74%. Results indicated that the WCST was consistently able to discriminate between patients and nonpatients, but was unable to discriminate among patient groups. These results suggest that, although the WCST measures executive abilities most commonly associated with the frontal lobe, it is not exclusively applicable to frontal lobe lesions.

In a recent study (Ozonoff 1995), both reliability and validity of the WCST were examined. The WCST was administered to three independent samples of autistic individuals. In Study 1, both autistic and learning disabled participants were administered the WCST in a test-retest design. In Study 2, both autistics and normal controls were administered both the computerized and traditional versions of the WCST to determine alternate format reliability. In the final project, Study 3, autistic and normal controls were administered the computerized version of the WCST to examine instrument validity.

In study one, 17 autistic and 17 learning disabled controls (matched on IQ, age and gender) were administered the traditional version of the WCST, as well as subtests information, vocabulary, block design and object assembly from the Weschler Intelligence Scale for Children – Revised (WISC-R) to obtain verbal, performance, and full scale IQ estimates. Approximately 2-1/2 years later, participants were recontacted for retesting. Results for both groups indicated high test-retest coefficients, with the autistic group receiving coefficients of .94 for total errors, and .93 for perseverative responses, and the learning disabled group receiving coefficients of .90 for total errors and .94 for perseverative responses. Ozonoff (1995) does caution that due to a lengthy between-test interval, the reliability coefficients may be elevated. With a shorter interval period, retesting may produce higher scores due to recognition and remembrance of the categorization rules.

In study two, 10 children and adolescents with autism, and 11 children and adolescent controls (matched on IQ and age) were administered the computerized version of the WCST at time 1 and the standard, traditional version of the WCST at time 2 (approximately one year later). Low to moderate alternative format reliability was found in both the autistic group (total errors, *r* = .49; perseverative responses, *r* = .73) and the control group (total errors, *r* = .60, perseverative responses *r* = .42). With only the perseverative responses (*r* = .73) for the autistic group falling within an adequate range, the results indicate that the two versions of the WCST may not be comparable.

Finally, in study three, 24 autistic children and adolescents and 24 normal controls (matched on age) were administered either the computerized WCST or the standard, traditional WCST. The groups were randomly split, with 12 autistic and 12 controls completing the computerized version and 12 autistic and 12 controls completing the traditional version. Both between-groups and within-group analyses were conducted, and means were obtained for total errors, perseverative responses, and number of categories completed. Results indicated that the autistic group performed poorer than the control group in all cases of standard administration, but no significant between-group differences were noted with computerized administration. Within-group differences did not reach significance for either group, although the autistic group performed better on the computer version.

Weschler Adult Intelligence Scale-Revised, Neuropsychological Instrument (WAIS-R NI)

The WAIS-R NI contains new subtests and modifications of the Wechsler Adult Intelligence Scale-Revised® (WAIS-R®; Psychological Corporation, San Antonio, TX, www.psychcorp.com), that permit a finer analysis of test behavior and provide profiles of spared and impaired cognitive functions. The new subtests are Sentence Arrangement, Spatial Span, and Symbol Copy which tap additional cognitive functions not assessed with the WAIS-R. Additionally, multiple-choice versions of the Information, Vocabulary, and Similarities subtests are included. This format allows for group testing, and benefits individual participants who provide brief answers or have difficulty accessing stored information (Psychological Corporation, 1991). Additionally, Kaplan, Fein, Morris, and Delis (1991) suggested that the multiple choice subtests expose intelligence or problem-solving abilities that may not be obvious in an individual=s response to a free-style format. They further contend that such a format should provide for better performance by participants. However, they did speculate that some participants may perform more poorly on multiple-choice items due to either choices sounding similar, or to being Apulled to opposites@ (p. 9). These multiple-choice versions were utilized in the current research. No research relevant to malingering is available for this instrument in a malingering-type study. WAIS-R NI subtests measure floor effect, performance curve, magnitude of error, and symptom validity.

Reliability and validity of the WAIS-R NI. The one study located that examined the WAIS-R NI expressed concerns regarding its utility as a clinical instrument. As discussed by Slick, Hopp, Strauss, Fox, Pinch and Stickgold (1996) Anormative data is lacking for the new...variables, imposing considerable limits on the WAIS-R NI=s clinical usefulness@ (p. 123). Therefore, they examined the test-retest reliability of the WAIS-R NI as compared to the WAIS-R. They administered the WAIS-R NI Information, Vocabulary, Similarities, Picture Arrangement, Block Design, Object Assembly and Digit Symbol subtests to a group of 20 Caucasian adults (10 male, 10 female). Both the standard subtest and the multiple choice versions were given for Information, Vocabulary and Similarities. Testing performance was carefully monitored and recorded so that scaled scores could be calculated that were as similar as possible to scoring criteria in the WAIS-R manual. The same participants were retested 19-29 days later using the administration instructions from the WAIS-R manual on the second administration. Correlations between the two testing sessions were calculated for each subtest administered, and the following coefficients were obtained: Information (r = .81), Vocabulary (r = .81), Similarities (r = .56), Picture Arrangement (r = .70), Block Design (r = .57), Object Assembly (r = .33), and Digit Symbol (r = .91). The test/retest scores were then compared to those reported by Wechsler in 1981 using univariate *t*-tests. The participants in the Slick et al., (1996) study showed a significantly greater gain in scaled scores, with the exception of Vocabulary scores. This gain was paralleled by Wechsler=s normative participants. Examination of the multiple choice versions of the subtests indicate both consistent, and poorer performance. Specifically, they scored similarly between multiple-choice versions of Information (*M* raw score = 22.1, *S.D.* = 3.3) and Similarities (*M* raw score = 26.1, *S.D.* = 1.8) and the standard versions (Information *M* raw score = 20.8, *S.D.* = 4.5; Similarities *M* raw score = 23.9, *S.D.* = 1.8). However, on the multiple choice version of Vocabulary, 30-85% of the participants obtained lower scores on 15 of 35 items (43%; *M* = 52.2, *S.D.* = 5.4) than they did on the standard version of the subtest (*M* = 56.3, *S.D.* = 6.9). The authors conclude that, with the exception of the variable performance on multiple choice versions, the scaled scores and learning effects at retest were consistent between the WAIS-R NI and the original WAIS-R.

Procedures

The participants belonged to one of the five previously described groups, and each were administered the six assessment measures in the specified order. Inclusion of neurologically-impaired patients was based on the diagnosis of head injury, which was determined by the owner of Neuropsychology Associates of Dallas, a licensed clinical psychologist in private practice. Brain-injured participants were then assigned to either group three (head injury with litigation) or group four (head injury without litigation). Litigation involvement was a direct result of the situation that allegedly caused the head injury. As noted by Iverson and Tulsky (2003) relevant comparison groups are necessary in malingering research, specifically litigating patients with brain injuries and nonlitigating patients with brain injury.

An individual with litigation that has no relationship to the head injury being assessed was not considered for the current research. The control, simulating, and forensic participants were asked to participate in exchange for a financial reward, and were required to provide informed consent (Appendixes A and B). The simulating group received precise instructions for simulation (Appendix C) that emphasized the feigning of a head injury (e.g., try to produce the most severe cognitive disabilities that you can without making it obvious).

Data was collected by advanced psychology doctoral students from the University of North Texas, specifically the primary researcher, and graduate student research volunteers with backgrounds and training in neuropsychology, specifically a graduate-level class in neuropsychological assessment, and hands-on experience administering neuropsychological tests to patient populations. As an additional control, the researchers in this study were blind to the membership of the control and simulating groups both prior to, and during, the administration of the neuropsychological instruments.

Due to the involvement of human participants in this study, approval was obtained with the University of North Texas Institutional Review Board for the Protection of Human Subjects in Research (IRB) through the Office of Research Administration and Academic Grants. Approval was provided by William Bruce Jones, Ph.D. (Owner, Neuropsychology Associates of Dallas) for inclusion of our assessment instruments in his standardized neuropsychological battery; subsequently, data was gathered from his head-injured patient population for groups three and four. Approval was also obtained from the Volunteers of America Ohio River Valley, Inc., for group five participants from their forensic population. To protect confidentiality of all participants, code numbers were assigned to each participant=s test results, so that no identifying information could ever be released.

Following the administration of the test battery, the simulating (Appendix E), control, and forensic participants (Appendix F) were asked to complete a post-assessment questionnaire to examine: (a) participants' understanding of the instructions, (b) how hard participants tried to follow the instructions, and (c) participants' knowledge of brain functioning and brain injury. Additional questions were included for the simulating group, including: (a) how successful participants felt they were in malingering believable brain damage, (b) how successful participants felt they were in deceiving the examiner, and (c) the specific strategies participants utilized in the malingering process. Questionnaires included both Likert scale and brief answer formats, and provided information regarding participant compliance and malingering sophistication. Following the questionnaire, a debriefing was provided.

Statistical Testing of the Research Hypotheses

To statistically test the following three research hypotheses, multivariate analysis of variance (MANOVA) were computed for each of the respective hypotheses. Since ANOVA testing with multiple correlated dependent measures can lead to inflation of type I error rate, MANOVA was performed to safeguard against overall type I error inflation (Tabachnick and Fidell, 2001). The multivariate statistical null hypothesis predicted no difference in group means for the optimally weighted composite of all the assessment instruments. Given a statistically significant multivariate F test (a rejected null hypothesis), follow-up ANOVAs were performed for each dependent measure. For each statistically significant ANOVA where there were more than two groups, a post-hoc analysis on pairwise means was performed.

Hypothesis One

The first hypothesis was that participants who attempted to malinger would produce lower scores on the RMT, the COWA, the Finger Oscillation test, and the WAIS-R NI than head-injured without litigation and control participants, yet would produce scores similar to head-injured with litigation and forensic participants. Likewise, it was hypothesized that subjects who attempted to malinger will produce higher scores on the WCST and the BCT than head-injured without litigation and control participants, yet would produce scores similar to head-injured with litigation and forensic participants.

To statistically test hypothesis one, a MANOVA was computed on the mean scores for each assessment instruments for each group. For each statistically significant ANOVA, post-hoc analyses on pairwise means were performed.

Additionally, based on the raw and T-scores obtained, cutting scores were determined to assist in the future identification of malingerers. Determination of cutting scores was based the percentages obtained for the concepts of sensitivity, specificity, PPV, and NPV as defined previously (Macmillan and Creelman, 1991). The formulas that were used to obtain sensitivity, specificity, PPV, & NPV are as follows: (1) sensitivity = true positives/(true positives + false negatives), (2) specificity = true negatives/(true negatives + false positives), (3) PPV = true positives/(true positives + false positives), and (4) NPV = true negatives/(true negatives + false negatives).

Receiver Operating Characteristic (ROC) curves were also utilized as a guideline in determining cutting scores. When examining likelihood data, sensitivity is favored over specificity, but not overly so. Therefore, a specificity percentile below 50 is undesirable. Accuracy is measured by the area under the ROC curve; the greater the area, the greater the diagnostic accuracy. An area of 1 represents a perfect test; while an area between .8 to .9 is considered good, between .7 and .8 is considered fair, and under .7 is considered poor.

Hypothesis Two

The second hypothesis was that the magnitude of error for each incorrect response (the degree of wrongness) for each participant could be an indication of the level of malingering or motivation to malinger, with the simulating group producing a greater degree of wrongness. To test this hypothesis, an average magnitude of error per incorrect answer was calculated for the WAIS-R NI Vocabulary, Information, and Similarities subtests, the BCT, and the WCST. These scores were analyzed between groups using a MANOVA and followed by an ANOVA with multiple comparisons. For the WAIS-R NI, the average magnitude of error per incorrect answer was determined by utilizing the scoring system provided in the WAIS-R NI manual. The system specifically identified 0-point, 1-point, and 2-point answers, as well as phonetically incorrect answers, which were also identified as 0-point answers. For the BCT, the average magnitude of error per incorrect answer was determined by analyzing responses to each item and coding them into 2-point, 1-point, and 0-point answers. In addition to the correct response (2-point answer), responses based on similar but incorrect strategies (1-point answer), and responses based on irrelevant or non-existent strategies (0-point answer) were determined. Subtests one through six were utilized for this hypothesis. Subtest seven was excluded as it is an encapsulation of the previous six subtests and has no inherently novel and consistent pattern or strategy.

The same procedure utilized for the BCT was also utilized for the WCST. The average magnitude of error per incorrect answer was determined by coding correct responses into 2-point answers, responses which contain an incorrect but possible match into 1-point answers, and responses which clearly contain no match to any category or dimension into 0-point answers. After all answers had been entered into the database, they were then recoded as follows: 2-point answers became 0-point answers, 1-point answers remained 1-point answers, and 0-point answers became 2-point answers. This recoding strategy enabled the highest value for the most incorrect answer, thereby the greater the total magnitude of error, the higher the total score per individual per assessment instrument. Statistical analyses were then performed on the recoded answers.

Hypothesis Three

The third hypothesis was that the rate of decay, or performance curve, on the COWA, Finger Oscillation, and WAIS-R NI Vocabulary, Information, and Similarities subtests for the simulating group would be less than the rate of decay for both the head-injured without litigation and control groups, yet similar to the rate of decay for both the head-injured with litigation and forensic groups. This research hypothesis stated that participants who attempt to malinger tend to have a lower rate of decay, or less acute performance curve, on assessment instruments. It was speculated that the performance curve of the simulating group would be flatter or linear, as simulators are believed to not only miss earlier, easier items, but to also have difficulty decreasing their performance with increased item difficulty, showing more variability within trails or tasks. In contrast, honest responders would tend to display a decline in performance, seen as a downward curve, due to a decrease in ability with increasing item difficulty, commonly displaying less to no variability within trails or tasks. To statistically test the rate of decay between groups, a repeated measures MANOVA analysis with polynomial contrasts was conducted for the first third, second third, and last third mean scores for each participant. The difference in group slopes was determined by comparing the simulating group with the remaining four groups.

Hypothesis Four

The final hypothesis examined the efficacy of Bolter=s 14-item Performance Validity Index (VI) from his 1992 study in detecting malingering. The revised BCT answer sheet illustrating the VI will be examined to assess its viability in successfully discriminating between the simulating group and all other groups. It is hypothesized that the simulating group would obtain more errors overall on the BCT, and therefore, would be more likely to obtain a higher number of VI errors than the control, head-injured with litigation, head-injured without litigation, and forensic groups. To statistically test this hypothesis, a one-way ANOVA with multiple comparisons was conducted using groups 1-5 as the grouping variable and total IV errors as the dependent variable.

CHAPTER III

RESULTS

Sample Characteristics

The sample was comprised of 126 participants, with five groups of 25 and 26 participants each (see Table 1 and Table 2).

Table 1: *Participants' Age by Group*

	Mean	S.D.	Minimum	Maximum
CP	44.80	12.41	26	70
SP	36.44	12.89	20	69
HL	41.24	13.31	18	69
HN	43.44	14.33	18	72
FP[a]	33.77	7.04	18	46
Total	39.89	12.76	18	72

NOTE: $F_{(df=4)} = 3.735$, $p = .007$. [a]FP mean age is significantly lower than the mean ages for CP ($p = .011$) and HN ($p = .038$). CP = Control Participants, SP = Simulating Participants, HL = Head-Injured with Litigation, HN = Head-Injured without Litigation, and FP=Forensic Participants.

Group one consisted of 25 control participants (CP) who were asked to respond honestly. Twelve participants were male (48%), and 13 were female (52%). The racial composition of the group was 19 Caucasian Americans (76%), 4 African Americans (16%), 1 Hispanic American (4%) and 1 Native American (4%). The participants had a mean age of 44.80 years (*S.D.* = 12.41) and an average of 13.80 years of education (*S.D.* = 2.38). Twenty-three participants were right hand dominant (92%).

Group two consisted of 25 simulating participants (SP) who were asked to feign a Abelievable@ head injury from an automobile accident. Malingering instructions for the simulation group can be found in Appendix C. Eleven participants were male (44%), and 14 were female (56%). The racial composition of the group was 16 Caucasian Americans (64%), 7 African Americans (28%), and 2 Hispanic Americans (8%). The participants had a mean age of 36.44

years (*S.D.* = 12.89) and an average of 12.32 years of education (*S.D.* = 2.75). All participants were right hand dominant (100%). Participants for the control and simulating groups were volunteers from the community who were randomly assigned to one of the two groups via instructions provided in a sealed envelope.

Group three consisted of 25 patients diagnosed with mild to mild/moderate head injuries who, due to their head injury, were involved in litigation (HL; Head Injured with Litigation) at the time of assessment. Fifteen participants were male (60%), and 10 were female (40%). The racial composition of the group was 19 Caucasian Americans (76%), 4 African Americans (16%), and 2 Hispanic Americans (8%). The participants had a mean age of 41.24 years (*S.D.* = 13.31) and an average of 12.84 years of education (*S.D.* = 1.72). Twenty-four participants were right hand dominant (96%). Group four consisted of 25 patients diagnosed with mild to mild/moderate head injuries who were not involved in litigation (HN; Head Injured with no litigation) at the time of assessment. Twenty-two participants were male (88%), and 3 were female (12%). The racial composition of the group was 23 Caucasian Americans (92%) and 2 African Americans (8%). The participants had a mean age of 43.44 years (*S.D.* = 14.33) and an average of 13.72 years of education (*S.D.* = 2.65). All participants were right hand dominant (100%). Groups three and four were included to provide profiles depictive of a neurologically impaired sample of the population for comparison purposes.

Finally, group five consisted of 26 adult males on parole for felony offenses (FP; Felony Parolees). The racial composition of the group was 9 Caucasian Americans (34.6%) and 17 African Americans (65.4%). The participants had a mean age of 33.73 years (*S.D.* = 7.13) and an average of 10.73 years of education (*S.D.* = 1.97). Twenty-two participants were right hand dominant (84.6%), 2 were left hand dominant (7.7%) and 2 were ambidextrous (7.7%). As mentioned previously, the inclusion of this group is based mainly on the premise that malingering (as well as head-injury) is high in the forensic population (Rogers, 1986), and will therefore permit a comparison between perceived malingerers, simulators, and the head-injury groups.

Table 2: *Participants' Gender, Ethnicity, Education and Handedness by Group*

	CP	SP	HL	HN	FP	Total	%ile	Total N
Gender								
Male	12	11	15	22	26	86	68%	
Female	13	14	10	3	0	40	32%	126
Ethnicity								
Caucasian 19	16	19	23	9	86	68%		
African-Am	4	7	4	2	17	34	27%	
Hispanic-Am	1	2	2	0	0	5	4%	
Asian-Am	0	0	0	0	0	0	0%	
Other	1	0	0	0	0	1	2%	126
Education								
Less than 9	0	2	0	0	4	6	5%	
9 to 11	2	7	3	2	13	27	21%	
12	7	4	12	11	6	40	32%	
13 to 15	9	8	7	5	2	31	25%	
16/Bachelor's	4	3	3	5	1	16	13%	
Graduate	3	1	0	2	0	6	5%	126
Handedness								
Right	23	25	24	25	22	119	94%	
Left	2	0	1	0	2	5	4%	
Ambidextrous	0	0	0	0	2	2	2%	126

NOTE: CP = Control Participants, SP = Simulating Participants, HL = Head-Injured with Litigation, HN = Head-Injured without Litigation, and FP = Forensic Participants.

Between-Group Differences on Demographics

Between-groups differences were found for age ($F_{(4)} = 3.735$, $p = .007$, $\eta^2 = .110$, $p = .001$, $1 - \beta = .876$; see Table 1), gender ($X^2_{(4)} = 28.898$, $p = .000$, w = .479), ethnicity ($X^2_{(12)} = 34.476$, w = .525), education ($X^2_{(40)} = 65.695$, $p = .006$, w = .721), and estimated IQ ($F_{(4)} = 10.634$, $p = .000$, $\eta^2 = .260$, $1 - \beta = 1.000$; see Table 3), but not for handedness ($X^2_{(8)}$, 2.061, $p = .149$, w = .309).

Table 3: *Participants' Estimated IQ by Group*

	Mean	S.D.	Minimum	Maximum
CP	98.80	8.69	85	120
SP	87.16	8.03	75	105
HL	93.60	6.54	85	110
HN	96.52	9.93	79	124
FP[a]	86.81	9.31	79	115
Grand Mean	92.53	9.58	75	124

NOTE: $F_{(df=4)} = 10.634$, $p = .000$. [a]FP and SP estimated IQ scores were significantly lower than the estimated IQs for the CP (p=.000) and HN (p=.003 and p=.005, respectively) groups. CP = Control Participants, SP = Simulating Participants, HL = Head-Injured with Litigation, HN = Head-Injured without Litigation, and FP = Forensic Participants.

The average age of the participants in this study was 39.9 years (Range: 18-72 years). The FP group was the youngest with a mean age of 33.8 (Range: 18-46), the SP group fell in the middle with a mean age of 36.4 (Range: 18-69), while the CP, HN, and HL groups were the oldest (*M* ages of 44.80, 43.44, and 41.24 respectively). Both CP and HN groups were significantly older than both the FP (p_= .002 and p = .006, respectively) and the SP (p = .017 and p = .045, respectively) groups. The FP group was also significantly younger than the HL group (p = .031).

Of the 126 participants, 86 were male and 40 were female. The CP group had almost an equal ratio of males (48%) to females (52%), and the SP group only slightly less so with 44% males and 56% females. The HL group had more males (60%) than females (40%). The HN group had considerably more males (88%) than females (12%), while the FP group was 100% males. The latter two groups were statistically different from the CP, SP, and HL groups. Caucasians comprised 68% of the sample, with African-Americans the second largest group (27%). Hispanic- and Native-Americans made up the remainder of the sample with less than 5% each.

The majority of the participants graduated from high school (73%). 25% had some college past high school, while 17% had a bachelor=s degree, and 5% had at least one year of graduate education. 21% of the sample had some high school education, but had not graduated, while 5% had only a junior high education. The average estimated mean IQ score across groups was 92.5 (Range = 75-124). Sixteen estimated IQs were Borderline (13%; Range = 70-79), 27 estimated IQs were Low Average (20%; Range = 80-89), 75 estimated IQs were Average (60%; Range = 90-109), six estimated IQs were High Average (5%; Range = 110-119), and two estimated IQs were Superior (2%; Range =120-129). The mean IQ for the two head-injured groups did not differ from each other (HL = 93.6; HN = 96.5), nor did they differ significantly from the control group (*M* = 98.8). Both the FP (*M* = 86.81) and SP (*M* = 87.16) groups, however, did have significantly lower mean IQ scores than the other three groups, with p values ranging from .000 to .007. The Tukey HSD post-hoc analysis yielded two homogeneous subsets: 1) the SP and FP groups, and 2) the CP, HL, and HN groups.

Right-handers (94%) were by far the largest group in the sample, with left-handers and participants who are ambidextrous each comprising less than 5% of the sample. The findings for handedness are grossly representative of the population in general.

Hypothesis One Inferential Statistics

Hypothesis one predicted that participants who attempted to malinger would produce poorer scores on the Rey-15 Item Memory Test (RMT; Andre Rey, 1964), the Controlled Oral Word Association Test (COWA; Bechtoldt, Benton, & Fogel, 1962), the HRNB Finger Oscillation Test (Halstead & Wolfson, 1993), and the Wechsler Adult Intelligence Scale-Revised® Neuropsychological Instrument Vocabulary, Information, and Similarities subtests (WAIS-R® NI; Psychological Corporation, San Antonio, TX, www.psychcorp.com) than the head-injured without litigation and control participants, yet produce scores similar to the head-injured with litigation and forensic participants. Likewise, it was hypothesized that subjects who attempt to malinger would produce more errors on the HRNB Booklet Category Test (Halstead & Wolfson, 1993) and Wisconsin Card Sorting Test^{TM} (WCST^{TM}; Psychological Assessment Resources, Lutz, FL, www.parinc.com) than the head-injured without litigation and control participants, yet produce scores similar to the head-injured with litigation and forensic participants.

To statistically test this hypothesis, a MANOVA was computed on the mean scores for the two groups (simulating/forensic/head-injured with litigation group, compared to the control/head-injured without litigation group) on the set of assessment instruments, followed by an ANOVA for each assessment instrument (see Table 4). The multivariate F statistic, Wilk's Lambda, was statistically significant for the main effect of group: $\lambda = .70$, $F_{(9)} = 5.231$, $p = .000$, $\eta^2 = .30$, $1-\beta = .999$.

Table 4: *Descriptive Statistics for Assessment Instruments by Malingering Versus Nonmalingering Groups*

Assessment Instrument	Group	N	Mean	Min to Max	S.D.	Variance
WAIS-R NI Vocabulary	1	50	48.22	36 to 66	6.27	39.277
	2	76	42.59	33 to 60	5.39	29.071
WAIS-R NI Information	1	50	53.76	29 to 70	8.135	66.186
	2	76	45.11	29 to 63	8.252	68.095
WAIS-R NI Similarities	1	50	62.18	40 to 81	8.889	79.008
	2	76	53.72	33 to 79	12.509	156.469

(*table continues*)

Table 4 (*continued*)

Assessment Instrument	Group	N	Mean	Min to Max	S.D.	Variance
RMT Total Items	1	50	13.22	6 to 15	2.566	6.583
	2	76	11.96	8 to 15	2.013	4.051
RMT Total Groups	1	50	4.26	2 to 5	.986	.972
	2	76	4.01	2 to 5	.887	.786
COWA Total Raw Words	1	50	32.66	8 to 52	11.162	124.584
	2	76	30.35	5 to 56	9.73	94.670
Finger Tapping Test	1	50	43.291	21 to 65	10.676	40.656
Dominant Hand	2	76	34.295	23 to 60	9.440	39.710
Booklet Category Test	1	50	55.78	10 to 160	32.953	1085.930
	2	76	58.68	13 to 109	31.066	965.072
Wisconsin Card Sorting	1	49	23.85	7 to 83	19.840	397.029
	2	76	40.56	5 to 104	28.283	823.877

NOTE: Group 1 = Control (CP) and Head-Injured without Litigation (HN) Groups; Group 2 = Simulating (SP), Head-Injured with Litigation (HL), and Forensic (FP) Groups.

A statistically significant difference was found for the main effect of group on the following assessment instruments: WAIS-R NI Vocabulary T-scores ($F_{(1)} = 28.224$, $p = .000$, $\eta^2 = .193$, $1 - \beta = 1.00$); Information T-scores ($F_{(1)} = 29.896$, $p = .000$, $\eta^2 = .202$, $1 - \beta = 1.00$); and Similarities T-scores ($F_{(1)} = 18.322$, $p = .000$, $\eta^2 = .134$, $1 - \beta = .989$); as well as the total number of words generated on the COWA ($F_{(1)} = 3.515$, $p = .063$, $\eta^2 = .029$, $1 - \beta = .460$); and the total number of perseverative errors on the WCST ($F_{(1)} = 14.749$, $p = .000$, $\eta^2 = .111$, $1 - \beta = .968$). However, not all the expected differences were found, therefore the hypothesis was not supported.

Table 5: *Descriptive Statistics for Assessment Instruments by All Groups*

Assessment Instrument	Group	N	Mean	Min to Max	S.D.	Variance
WAIS-R NI Vocabulary T-Score	CP	25	49.00	40 to 63	5.90	34.833
	SP	25	41.28	33 to 53	5.24	27.460
	HL	25	45.48	40 to 57	4.42	19.510
	HN	25	47.44	36 to 66	6.64	44.090
	FP	26	41.08	36 to 60	5.45	29.674
Group Mean		126	44.83	33 to 66	6.36	40.481
WAIS-R NI Information T-Score	CP	25	53.56	29 to 70	7.94	63.007
	SP	25	42.28	33 to 60	7.59	57.627
	HL	25	50.32	36 to 63	7.72	59.560
	HN	25	53.96	36 to 70	8.48	72.040
	FP	26	42.81	29 to 60	7.15	51.122
Group Mean		126	48.54	29 to 70	9.21	84.874
WAIS-R NI Similarities T-Score	CP	25	63.12	40 to 81	9.96	99.110
	SP	25	45.76	33 to 79	10.90	118.857
	HL	25	63.12	46 to 79	8.37	70.110
	HN	25	61.24	46 to 79	7.77	60.357
	FP	26	52.35	36 to 79	11.59	134.315
Group Mean		126	57.08	33 to 81	11.921	142.106
RMT Total Items	CP	25	13.12	9 to 15	2.37	5.610
	SP	25	11.96	9 to 15	1.70	2.873
	HL	25	14.16	11 to 15	1.25	1.557
	HN	25	13.32	6 to 15	2.79	7.810
	FP	26	12.73	8 to 15	2.32	5.405
Group Mean		126	13.06	6 to 15	2.24	5.029
RMT Total Groups	CP	25	4.20	2 to 5	.96	.917
	SP	25	3.48	3 to 5	.77	.593
	HL	25	4.56	3 to 5	.58	.340
	HN	25	4.32	2 to 5	1.03	1.060
	FP	26	4.00	2 to 5	.94	.880
Group Mean		126	4.11	2 to 5	.93	.868
COWA Total Raw Words	CP	25	35.68	27 to 51	6.47	41.893
	SP	25	27.20	5 to 43	9.27	85.833
	HL	25	33.12	14 to 56	10.39	100.260
	HN	24	29.04	8 to 52	12.24	149.694
	FP	26	27.31	8 to 42	9.68	93.742
Group Mean		125	31.16	5 to 56	10.38	101.741

(*table continues*)

Table 5 (*continued*)

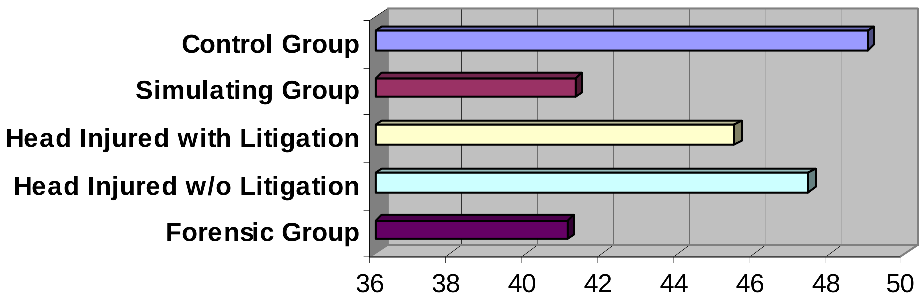
Assessment Instrument	Group	N	Mean	Min to Max	S.D.	Variance
Finger Oscillation Test Dominant Hand	CP	24	43.303	32 to 57	6.376	40.656
	SP	24	35.725	23 to 46	6.30	39.710
	HL	24	44.085	29 to 60	8.379	70.211
	HN	25	45.011	21 to 65	11.025	121.551
	FP	26	47.964	32 to 59	7.769	60.354
Group Mean		123	42.074	21 to 65	11.149	81.312
BCT Total Raw Errors	CP	24	48.92	23 to 91	20.89	436.775
	SP	25	50.00	20 to 90	21.31	454.250
	HL	24	65.58	20 to 160	40.41	1633.123
	HN	24	67.29	10 to 160	37.66	1418.737
	FP	26	62.92	13 to 109	26.06	679.114
Group Mean		123	57.48	10 to 160	31.85	948.143
BCT Total IV Raw Errors	CP	24	.25	0 to 3	.74	.543
	SP	25	.44	0 to 2	.71	.507
	HL	24	.79	0 to 9	2.26	5.129
	HN	24	1.04	0 to 6	1.81	3.259
	FP	26	.19	0 to 1	.40	.162
Group Mean		123	.54	0 to 9	1.39	1.923
WCST Total Errors	CP	25	23.68	8 to 93	19.784	391.393
	SP	25	55.44	8 to 104	31.641	1001.173
	HL	24	21.58	7 to 56	13.966	195.036
	HN	22	24.04	7 to 84	20.344	413.862
	FP	26	43.77	5 to 99	25.483	649.385
Group Mean		122	34.20	5 to 96	26.582	706.589

NOTE: CP = Control Participants, SP = Simulating Participants, HL = Head-Injured with Litigation, HN = Head-Injured without Litigation, and FP = Forensic Participants.

Further examination of the group means through group ANOVA analysis, and with the Student-Newman-Keuls (SNK) post-hoc test statistic revealed that the group differences were not all in the predicted direction (see Table 5).

On the Vocabulary subtest (see Figure 1), the CP group (*M* = 49) performed significantly better than both the SP (*M* = 41.28) and FP (*M* = 41.08) groups at the *p* = .000 level. The HN group

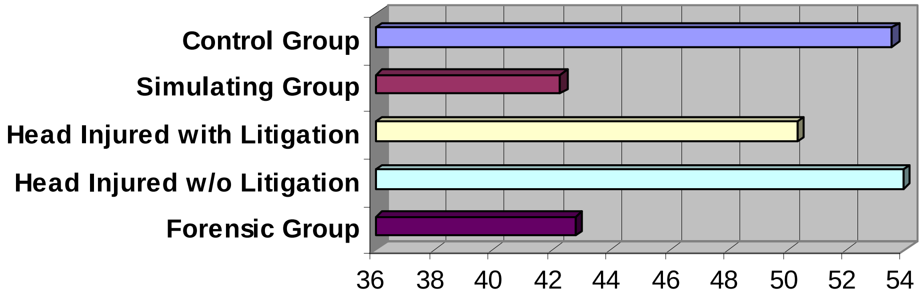
Figure 1: *Mean WAIS-R NI Vocabulary T-Score per Group*



(*M* = 47.44) also performed significantly better than both the SP (*p* = .008) and FP (*p* = .005) groups. The SP and FP groups were most alike (*p* = 1.000). The HL group (*M* = 45.48) was statistically different from the other four groups. The SNK post-hoc analysis yielded two homogeneous subsets: 1) the SP and FP groups, and 2) the CP, HL, and HN groups.

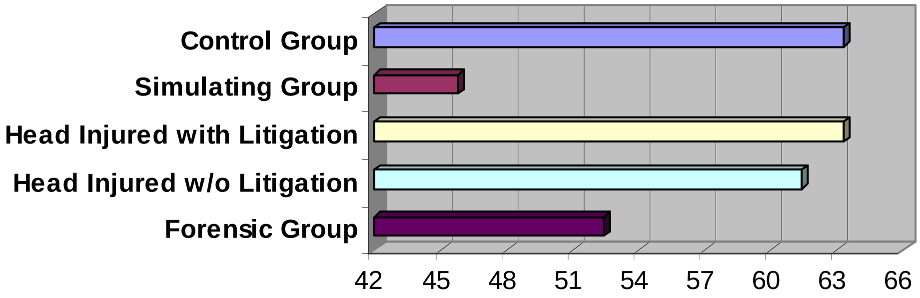
A similar pattern was found on the Information subtest (see Figure 2). Both the CP (*M* = 53.56) and HN (M = 53.96) groups performed significantly better than the SP (*M* = 42.28) and FP (*M* = 42.81) groups at the *p* = .000 level. The CP and HN groups were most alike (*p* = 1.000), as were the SP and FP groups (*p* = 1.000). The HL group (*M* = 50.32) performed significantly better than both the SP (*p* = .013) and FP (*p* = .023) groups. The SNK post-hoc analysis yielded two homogeneous subsets: 1) the SP and FP groups, and 2) the HL, CP, and HN groups. The Similarities subtest revealed a greater range between means (see Figure 3), and a reverse finding between the HL and HN groups than that originally hypothesized. The CP and HL groups (*M* =

Figure 2: *Mean WAIS-R NI Information T-Score per Group*



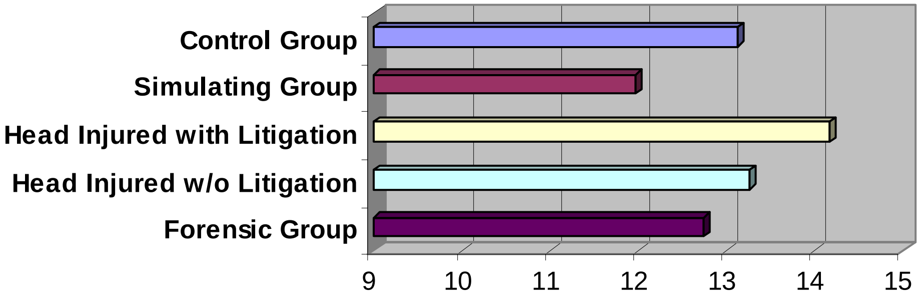
63.12; *p* = 1.000) performed statistically the same, and significantly better than the SP (*M* = 45.76; *p* = .000) and FP (*M* = 52.35; *p* = .006) groups. Similarly, the HN group (*M* = 60.5) also performed significantly better than the SP (*p* = .000) and FP (*p* = .039) groups. SNK post-hoc analysis yielded three homogeneous subsets: 1) the SP group, 2) the FP groups, and 3) CP, HL, and HN groups.

Figure 3: *Mean WAIS-R NI Similarities T-Score per Group*



Results for the RMT total items recalled (see Figure 4) were less remarkable, with only the SP (*M* = 11.96) and HL (*M* = 14.16) groups displaying significance

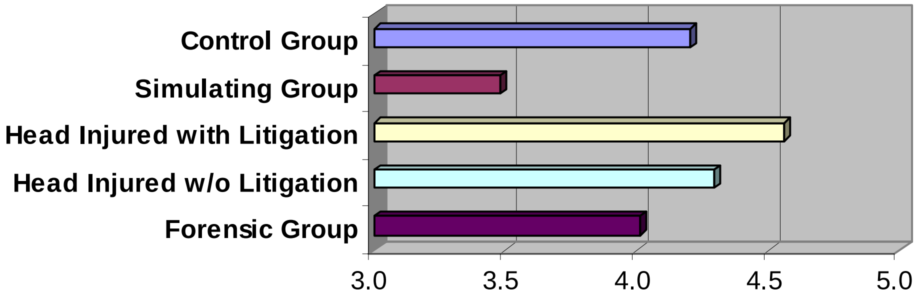
Figure 4: *Mean Total Number of RMT Items per Group*



with each other (*p* = .014). The remaining groups (FP, *M* = 12.73; CP, *M* = 13.12; and HN, *M* = 13.25) showed no significant difference from any other group. The SNK post-hoc analysis yielded two homogeneous subsets: 1) the CP, SP, HN, and FP groups, and 2) the HL, CP, HN, and FP groups.

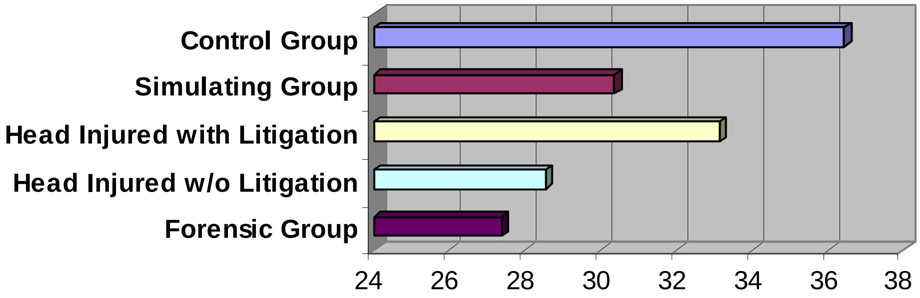
Findings for the RMT total categories obtained (see Figure 5) revealed a significantly poorer performance between the SP group, and both the HL (p = .001) and HN (p = .037) groups; however, no other group differences were noted. The SNK post-hoc analysis yielded two homogeneous subsets: 1) the SP group, and 2) the CP, HL, HN, and FP groups. These results are surprising given the nature of the RMT, and the research literature which, although somewhat equivocal, has displayed slightly stronger results for the support of the RMT as a screening measure for malingering. Our findings do not strongly support that contention.

Figure 5: *Mean Total Number of RMT Categories per Group*



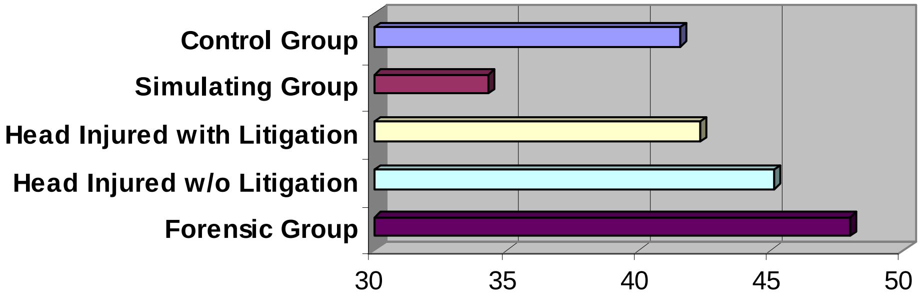
The number of words generated on the COWA did not result in any significant differences between groups (see Figure 6); however, the SP and FP groups performed identical (p = 1.000). The SNK post-hoc analysis yielded two homogeneous subsets: 1) the FP, SP, HL, and HN groups, and 2) the CP, SP, HL, and HN groups. The mean total words generated for each group was 36.44 (CP), 30.36 (SP), 33.12 (HL), 28.55 (HN), and 27.39 (FP), with the control group generating the most amount of words, and the forensic group generating the least amount. However, the simulating group generated more words than the head injured without litigation, therefore, the COWA does not appear to be an adequate indicator for malingering detection.

Figure 6: *Mean Total Number of COWA Words per Group*



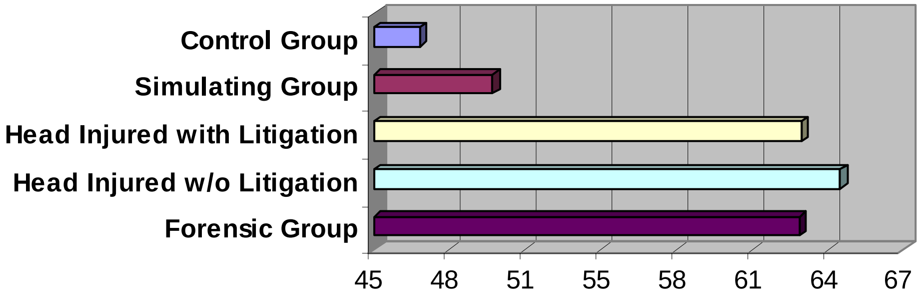
Results on dominant hand finger oscillation (see Figure 7) revealed a significantly poorer performance between the SP group (*M* = 35.725) and both the HN (*M* = 45.011; *p* = .12) and FP (*M* = 47.964; *p* = .00) groups. The SNK post-hoc analysis yielded two homogeneous subsets: 1) the SP group, and 2) the CP, HL, HN, and FP groups. The simulating group performed significantly poorer than all other groups (see Table 5). Interestingly, the forensic group performed the best on this instrument, obtaining the highest mean number of taps (*M* = 47.96).

Figure 7: *Mean Total Dominant Hand Finger Tapping Score per Group*



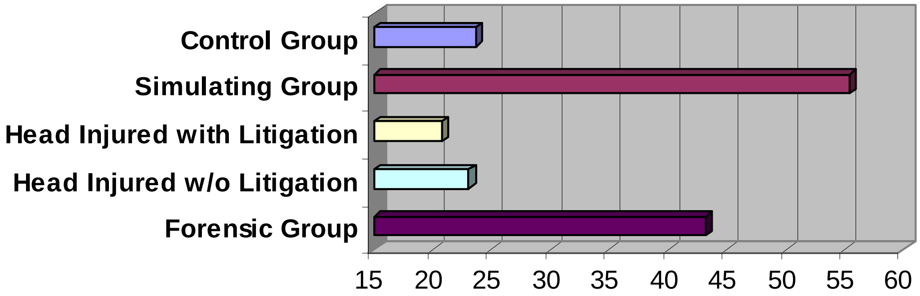
There were no significant between-group differences for total number of errors on the BCT ($F_{(4)}$ = 2.061, *p* = .090; see Figure 8), with the SNK post-hoc analysis yielding only one subset containing all groups. While there were no statistical differences, the CP (*M* = 48.92) and SP (*M* = 56.00) groups performed similar, while the HN (*M* = 65.58) and HL (*M* = 67.29) groups performed alike.

Figure 8: *Total Number of BCT Errors per Group*



Total number of errors obtained on the WCST (see Figure 9) revealed a statistically poorer performance (greater number of errors) between the SP group (*M* = 55.49), and the CP (*M* = 23.68; *p* = .000), HL (*M* = 21.58; *p* = 000), and HN (*M* = 24.04; *p* = .000) groups. The FP group (*M* = 43.77) was statistically poorer than the HL group (*p* = .023), and approached significance with the HN group (p = .060). The CP and HN groups performed most similar to each other (*p* = 1.000).

Figure 9: *Total Number of WCST Errors by Group*



SNK post-hoc analysis yielded two homogeneous subsets: 1) the CP, HL and HN groups, and 2) the SP and FP groups.

As a function of hypothesis one, cutting scores were determined, based on the percentages obtained for sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), using formulas outlined in the methods section. In addition, hit rates were calculated utilizing the total number of true positives, false positives, true negatives, and false negatives per cutoff score for each assessment instrument. Cutting score and efficacy tables 12-24, and ROC curve figure 17 - 25 (WAIS-R NI Vocabulary T-Score, p. 136**;** WAIS-R NI Information T-Score, p. 136; WAIS-R NI Similarities T-Score, p. 137; RMT Total Groups, p. 137; RMT Total Items, p. 138; COWA Total Words, p. 138; Total Finger Taps, p. 139; BCT Errors, p. 139; and WCST Errors, p. 140) can be found in the appendix. Table 18 reports a wide range of cutoff scores for all assessment instruments, along with the respective statistics, while individual instruments with select cutoff scores can be seen on Tables 12 through 17. The WAIS- R NI subtests evidenced the highest optimal hit rates (see Table 12), with 69.8% for Vocabulary (PPV = 60.7%; NPV = 77.1%), 66.7% for Information (PPV = 54.8%; NPV = 90.5%), and 59.2% for Similarities (PPV = 50.5%; NPV = 96.0%). Interestingly, although higher hit rates were obtained on the WAIS-R NI subtests, the cutoff scores fell in the average range, and reductions in NPV were also evidenced. As the purpose of the current study was to assist in the detection of malingering, hit rates with strong NPV percentiles would seem optimal. Less impressive results were obtained for the remaining instruments (see Table 13-24).

The relationship between the above statistics (specifically, true- and false-positives) is further illustrated in a receiver operating characteristic (ROC) curve (see Figure 17– 25). ROC curves are another factor utilized as a guideline in determining cutoff scores. When examining diagnostic data, sensitivity is favored over specificity, but not overly so; a decline of specificity below 50% is nondiagnostic. Accuracy is measured by the area under the ROC curve: An area of 1 represents a perfect test; while an area between .8 to .9 is good, between .7 and .8 is fair, and less than .7 is poor.

ROC analyses revealed fair findings for the WAIS-R NI Vocabulary subtest (Area = 0.756, *S.E.* = 0.046) and the WAIS-R NI Information subtest (Area = 0.788, *S.E.* = 0.043). Both the WAIS-R NI Similarities subtest (Area = 0.681, *S.E.* = 0.049), and more so the WCST total errors (Area = 0.694, *S.E.* = 0.047) approached fair results. Poor results were obtained for the RMT Total Items (Area = 0.588, *S.E.* = 0.053), RMT Total Groups (Area = 0.575, *S.E.* = 0.052), COWA total words (Area = 0.574, *S.E.* = 0.053), Finger Oscillation dominant hand (Area = 0.538, *S.E.* = 0.055), and BCT total errors (Area = 0.527, *S.E.* = 0.052).

A consideration for the current hypothesis is the speculation that the forensic group would be highly likely to malinger given the nature of its members, hence the forensic group was believed to perform similar to the simulating group. In reality, however, there were no incentives provided to the forensic group to alter their testing performance in the current study. Furthermore, the forensic group should not be confused with a “known-group” design, which allows for identification of known malingerers. A second consideration for the current study is that in practice, head-injured individuals represent a substantial number of the patients that present for neurocognitive evaluation for both insurance litigation and workman’s compensation claims, which in turn opens the possibly of malingering during such evaluations. Given this caveat in group identification, exploratory analyses were also conducted comparing the results for the simulating group against the results for both head-injured groups. Subtest means for the WAIS-R NI and Finger Oscillation, total raw scores for the COWA and RMT, and total errors for the BCT and WCST for each of the three groups were used to calculate cutting scores, based on the percentages obtained for sensitivity, specificity, PPV, and NPV, using formulas outlined in the methods section. In addition, hit rates were calculated utilizing the total number of true positives, false positives, true negatives, and false negatives per cutting score for each assessment instrument (See Table 12– 17).

Additional exploratory analyses were also conducted to examine the effects of education on the assessment instruments for both the total sample and for between-group differences (See Table 6 – 8). MANOVAs were conducted on the assessment instruments, using education as a covariate, followed by ANOVAs with multiple comparisons. A statistically significant difference was found for the effect of education for the entire sample on the following assessment instruments: WAIS-R NI Vocabulary T-scores ($F_{(1)}$ = 41.732, p = .000); WAIS-R NI Information T-scores ($F_{(1)}$ = 53.152, p = .000), and WAIS-R NI Similarities T-scores ($F_{(1)}$ = 28.777, p = .000); as well as the total number of words generated on the COWA ($F_{(1)}$ = 7.448, p = .007); and the total number of perseverative errors on the WCST ($F_{(1)}$ = 5.326, p = .023). The marginal means for the subtests for the total sample can be seen in Table 6.

Table 6: *Estimated Marginal Means for the Entire Sample with Education as a Covariate*

Assessment Instrument	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Vocabulary TScore	44.825^a	.492	43.851	45.800
Information TScore	48.540^a	.689	47.175	49.904
Similarities TScore	57.079^a	.961	55.178	58.981
Rey Total Items	13.056^a	.199	12.662	13.450
Rey Total Groups	4.111^a	.082	3.948	4.274
COWA Total Words	31.208^a	.899	29.429	32.987
Finger Tapping Mean	42.278^a	.993	40.314	44.243
Category Total Errors	57.532^a	2.826	51.939	63.124
WCST Total Errors	33.111^a	2.350	28.459	37.763

A statistically significant difference was also found for the effect of education by group on the following assessment instruments (See Table 7).

Table 7: *Between-Group Effects with Education as a Covariate on Assessment Instruments*

Assessment Instrument	F-Statistic	P Value	Partial Eta Squared	Observed Power
Education				
Vocabulary T-Score	21.362	.000	.151	.996
Information T-Score	29.712	.000	.198	1.000
Similarities T-Score	18.353	.000	.133	.989
RMT Total Items	.783	.378	.006	.142
RMT Total Groups	.856	.357	.007	.151
COWA Total Words	3.779	.054	.031	.487
Finger Oscillation Mean	1.297	.257	.011	.204
BCT Total Errors	.666	.416	.006	.128
WCST Total Errors	.512	.476	.004	.109
Group				
Vocabulary T-Score	5.554	.000	.156	.974
Information T-Score	7.910	.000	.209	.997
Similarities T-Score	12.548	.000	.295	1.000
RMT Total Items	3.171	.016	.096	.810
RMT Total Groups	4.977	.001	.142	.956
COWA Total Words	2.355	.058	.073	.667
Finger Oscillation Mean	6.518	.000	.178	.990
BCT Total Errors	1.649	.166	.052	.495
WCST Total Errors	9.501	.000	.241	1.000

Table 8: *Estimated Marginal Means by Group with Education as a Covariate on Assessment Instruments*

				95% Confidence Interval	
Assessment Instrument	**Group ID**	**Mean**	**Std. Error**	**Lower Bound**	**Upper Bound**
Vocabulary T-Score	**CP**	47.944[a]	1.057	45.851	50.037
	SP	41.603[a]	1.034	39.555	43.651
	HL	45.318[a]	1.033	43.274	47.363
	HN	46.458[a]	1.054	44.373	48.544
	FP	42.881[a]	1.085	40.733	45.028
Information T-Score	**CP**	51.871[a]	1.433	49.033	54.709
	SP	42.797[a]	1.403	40.019	45.574
	HL	50.062[a]	1.400	47.289	52.834
	HN	52.390[a]	1.429	49.561	55.219
	FP	45.693[a]	1.471	42.781	48.606
Similarities T-Score	**CP**	61.374[a]	1.885	57.641	65.107
	SP	46.294[a]	1.845	42.641	49.947
	HL	62.853[a]	1.842	59.206	66.500
	HN	59.617[a]	1.879	55.896	63.338
	FP	55.329[a]	1.935	51.498	59.159
Rey Total Items	**CP**	13.035[a]	.442	12.159	13.911
	SP	11.986[a]	.433	11.129	12.843
	HL	14.147[a]	.432	13.291	15.003
	HN	13.241[a]	.441	12.368	14.115
	FP	12.875[a]	.454	11.976	13.774
Rey Total Groups	**CP**	4.164[a]	.179	3.811	4.518
	SP	3.491[a]	.175	3.145	3.837
	HL	4.555[a]	.174	4.209	4.900
	HN	4.287[a]	.178	3.934	4.639
	FP	4.061[a]	.183	3.698	4.424
COWA Total Words	**CP**	35.590[a]	2.023	31.584	39.595
	SP	30.620[a]	1.980	26.700	34.540
	HL	32.990[a]	1.976	29.077	36.903
	HN	28.098[a]	2.017	24.105	32.090
	FP	28.837[a]	2.076	24.727	32.947
Finger Tapping Mean	**CP**	41.053[a]	2.103	36.890	45.216
	SP	34.454[a]	2.058	30.380	38.528
	HL	42.242[a]	2.054	38.175	46.309
	HN	44.529[a]	2.096	40.380	48.679
	FP	48.848[a]	2.158	44.576	53.120
Category Total Errors	**CP**	48.095[a]	6.430	35.364	60.825
	SP	49.653[a]	6.292	37.195	62.111
	HL	63.134[a]	6.281	50.697	75.570
	HN	65.655[a]	6.409	52.964	78.345
	FP	60.985[a]	6.598	47.921	74.049
WCST Total Errors	**CP**	24.421[a]	4.788	14.941	33.900
	SP	55.213[a]	4.685	45.937	64.490
	HL	20.833[a]	4.677	11.572	30.094
	HN	22.808[a]	4.773	13.359	32.258
	FP	41.927[a]	4.913	32.200	51.655

Hypothesis Two Inferential Statistics

The second research hypothesis states that the magnitude of error per each incorrect response on the WAIS-R NI Vocabulary, Information, and Similarities subtests, the BCT, and the WCST for the simulators should be greater than the magnitude of error per each incorrect response for the forensic, head- injured, and control groups.

To statistically test the difference in the magnitude of error per each incorrect response on the assessment instruments for the simulating group, the results for each participant were recoded following the procedures outlined in the methods section, with the purpose of identifying 0-point, 1-point, and 2-point answers for the assessment instruments under study. The recoded results were then analyzed using a MANOVA, followed by an ANOVA with multiple comparisons. The multivariate F statistic, Wilk's Lambda, was statistically significant for the main effect of group for the WAIS-R NI subtests ($\lambda = .515$, $F_{(12)} = 7.541$, $p = .000$, $\eta^2 = .198$, $1-\beta = 1.000$). Differences were found between-groups for the Vocabulary subtest ($F_{(4)} = 16.494$, $p = .000$, $\eta^2 = .351$, $1-\beta = 1.000$) (see Figure 10); Information subtest ($F_{(4)} = 9.413$, $p = .000$, $\eta^2 = .236$, $1-\beta = .999$) (see Figure 11); and Similarities subtest ($F_{(4)} = 15.927$, $p = .000$, $\eta^2 = .343$, $1-\beta = 1.000$ (see Figure 12).

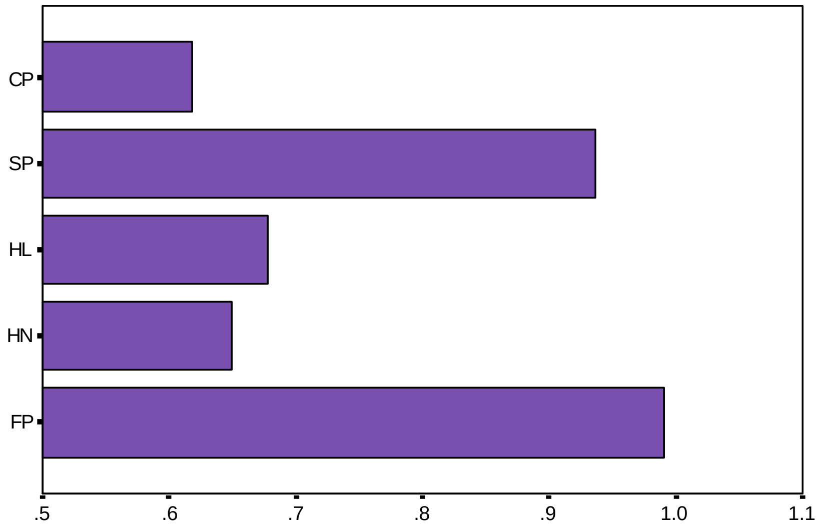
Further examination of the group means using the Scheffe post-hoc statistic revealed that the group differences were not all in the predicted direction (see Table 9). On the Vocabulary subtest, the SP group performed significantly poorer than the HN ($p = .000$),

Table 9: *Participants' Mean Magnitude of Error on the WAIS-R NI Subtests*

	Group	N	Mean	S. D.	Variance	Min	Max
Vocabulary Subtest	CP	25	.6183	.17591	.009	.26	1.03
	SP	25	.9371	.27529	.018	.49	1.43
	HL	25	.6777	.16711	.005	.40	.97
	HN	25	.6491	.22790	.061	.00	1.00
	FP	27	.9905	.22857	.015	.54	1.31
Total		127	.7780	.26716			
Information Subtest	CP	25	.4703	.20591	.008	.00	1.00
	SP	25	.7448	.23932	.012	.07	1.36
	HL	25	.5421	.19822	.006	.00	.64
	HN	25	.4524	.22815	.061	.00	1.00
	FP	27	.7011	.22388	.012	.07	1.43
Total		127	.5840	.24728			
Similarities Subtest	CP	25	.3086	.23933	.009	.21	1.21
	SP	25	.8314	.36356	.029	.34	1.14
	HL	25	.3114	.19765	.009	.21	1.00
	HN	25	.3600	.21673	.057	.00	.86
	FP	27	.6111	.37396	.036	.28	1.31
Total		127	.4865	.35186			

NOTE: CP = Control Participants, SP = Simulating Participants, HL = Head-Injured with Litigation, HN = Head-Injured without Litigation, and FP = Forensic Participants

Figure 10. *Magnitude of Error on the WAIS-R NI Vocabulary Subtest by Group*



On the Information subtest, the SP group again performed significantly poorer than the HN (*p* = .000), CP (*p* = .001), and HL (*p* = .036) groups. The CP and HN groups' performance were nearly identical (*p* = .999), followed by the CP and HL groups (*p* = .855). The SNK post-hoc analysis yielded two homogeneous subsets: 1) the CP, HL, and HN groups, and 2) the SP and FP groups.

Figure 11. *Magnitude of Error on the WAIS-R NI Information Subtest by Group*

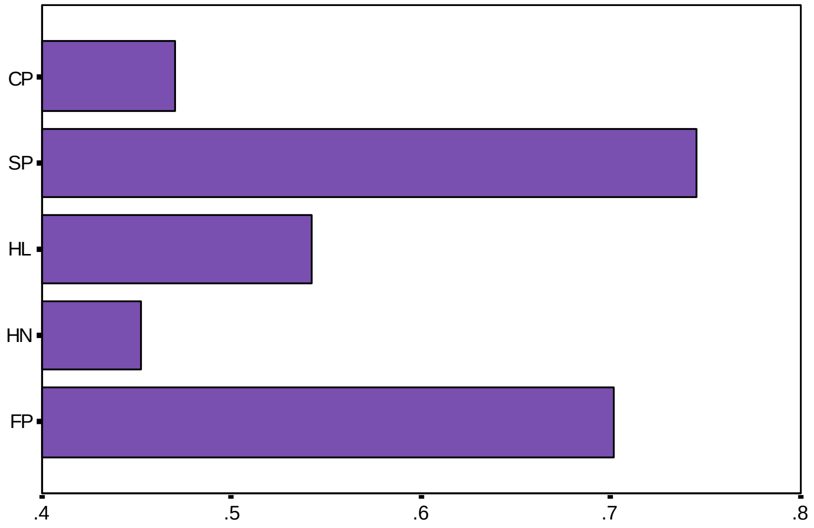
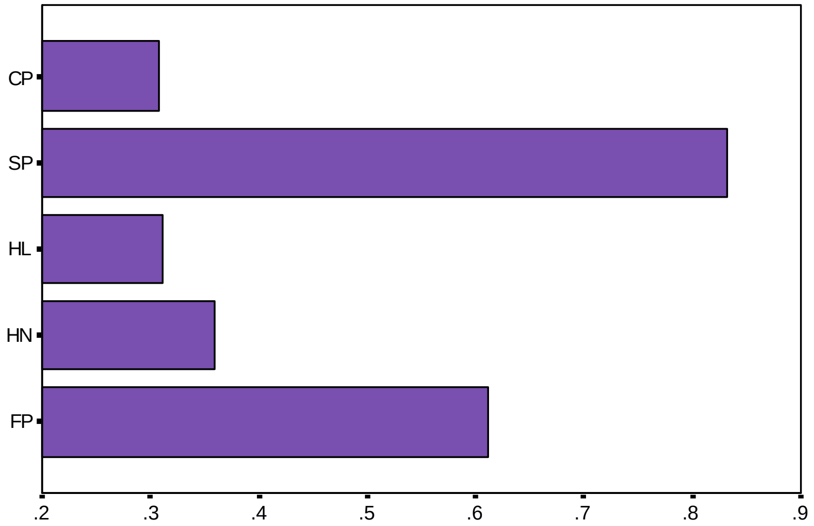


Figure 12. *Magnitude of Error on WAIS-R NI Similarities Subtest by Group*



The multivariate F statistic, Wilk's Lambda, was statistically significant for the main effect of group for the BCT subtests ($\lambda = .697$, $F_{(28)} = 1.565$, $p = .035$, $\eta^2 = .086$, $1 - \beta = .967$). Although a main effect for group was found on the BCT (see Table 10 and Figure 13), further examination of the group means through group ANOVA analysis revealed no significant differences between groups, and the SNK post-hoc test statistic revealed one homogenous subset for all five groups.

Table 10: *Participants' Mean Magnitude of Error on the BCT*

	Group	N	Mean	S. D.	Variance	Min	Max
Subtest 1	CP	25	.0000	.00000	.000	.00	.00
	SP	25	.0300	.10992	.012	.00	.50
	HL	25	.0400	.11815	.014	.00	.50
	HN	25	.0400	.09354	.009	.00	.25
	FP	27	.0000	.00000	.000	.00	.00
	Total	127	.0218	.08378	.007		
Subtest 2	CP	25	.0080	.01871	.000	.00	.05
	SP	25	.0320	.05752	.003	.00	.20
	HL	25	.0100	.02500	.001	.00	.10
	HN	25	.0020	.01000	.000	.00	.05
	FP	27	.0192	.05114	.003	.00	.25
	Total	127	.0143	.03840	.001		
Subtest 3	CP	25	.3900	.29545	.087	.00	.90
	SP	25	.4150	.23004	.053	.10	.85
	HL	25	.4690	.35328	.125	.00	.95
	HN	25	.5320	.34577	.120	.00	.93
	FP	27	.5221	.33078	.109	.00	.95
	Total	127	.4661	.31452	.099		
Subtest 4	CP	25	.3890	.30044	.090	.00	1.05
	SP	25	.4200	.26858	.072	.10	1.08
	HL	25	.4040	.36783	.135	.00	1.13
	HN	25	.3880	.39877	.159	.00	1.15
	FP	27	.5058	.32593	.106	.10	1.15
	Total	127	.4220	.33301	.111		

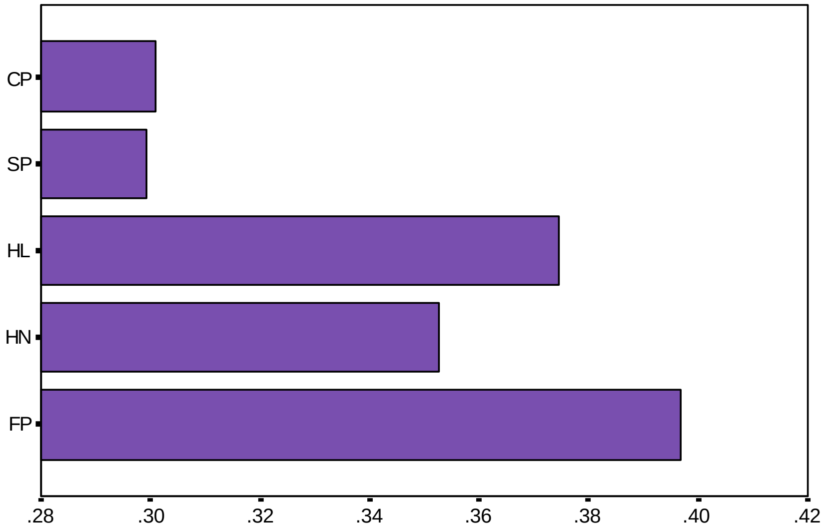
(*table continues*)

Table 10: (*continued*)

	Group	N	Mean	S. D.	Variance	Min	Max
Subtest 5	CP	25	.3700	.24367	.059	.00	1.00
	SP	25	.2920	.16421	.027	.13	.85
	HL	25	.4580	.30437	.093	.00	1.20
	HN	25	.4300	.34293	.118	.00	1.20
	FP	27	.4952	.20482	.042	.10	.95
	Total	127	.4097	.26550	.070		
Subtest 6	CP	25	.2530	.19913	.040	.00	.73
	SP	25	.2350	.12788	.016	.08	.53
	HL	25	.4030	.35593	.127	.00	1.25
	HN	25	.3200	.32420	.105	.00	1.38
	FP	27	.3317	.16622	.028	.10	.68
	Total	127	.3087	.25378	.064		
Total BCT Errors	CP	25	.3008	.14289	.020	.00	.56
	SP	25	.2992	.12726	.016	.13	.58
	HL	25	.3744	.21582	.047	.00	.93
	HN	25	.3527	.23704	.056	.00	.87
	FP	27	.3968	.16228	.026	.10	.73
	Total	127	.3452	.18325	.034		

NOTE: CP = Control Participants, SP = Simulating Participants, HL = Head-Injured with Litigation, HN = Head-Injured without Litigation, and FP = Forensic Participants.

Figure 13. *Magnitude of Error on the BCT by Group*



The multivariate F statistic, Wilk's Lambda, was statistically significant for the main effect of group for the WCST ($F_{(4)} = 9.022$, $p = .000$, $\eta^2 = .233$, $1 - \beta = .999$; see Table 11 and Figure 14). The SP group performed significantly poorer than the HN ($p = .000$), CP ($p = .001$), and HL ($p = .003$) groups, but not from the FP ($p = .603$) group. The FP group also performed significantly poorer than the HL group ($p = .038$). The CP and HN groups performed most alike ($p = .998$), followed by the CP and HL groups ($p = .997$). The SNK post-hoc analysis yielded two homogeneous subsets: 1) the CP, HL, and HN groups, and 2) the SP and FP groups.

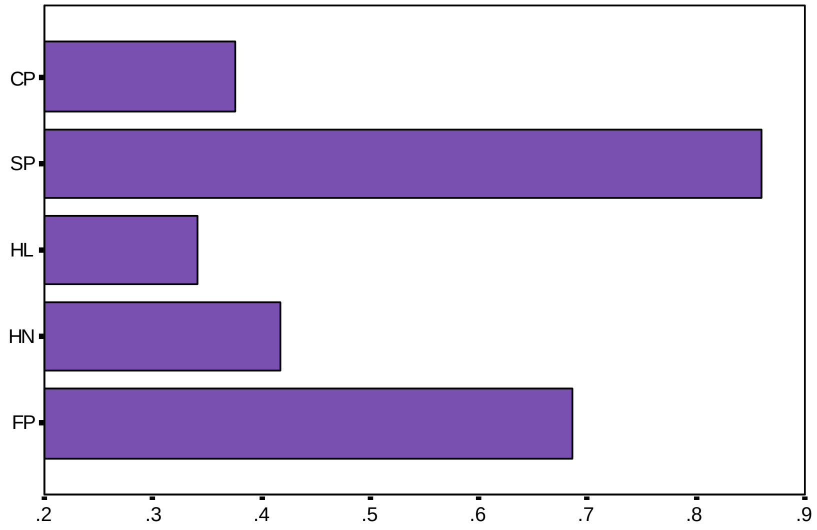
Although the magnitude of error theory adequately differentiated between the SP group, and the CP, HN and HL groups, it did not differentiate between the SP and FP groups, as was hypothesized. Therefore, the expected differences were not found, indicating that the hypothesis was not supported.

Table 11: *Participants' Mean Magnitude of Error on the WCST*

	Group	N	Mean	S. D.	Variance	Min	Max
Total Errors	CP	25	.37625	.294291	.087	.125	1.453
	SP	25	.86000	.493374	.243	.125	1.625
	HL	24	.34049	.215572	.046	.109	.875
	HN	24	.41797	.407977	.166	.109	2.000
	FP	26	.68570	.399153	.159	.078	1.578
Total		124	.53982	.421917	.178	.078	2.000

NOTE: CP = Control Participants, SP = Simulating Participants, HL = Head-Injured with Litigation, HN = Head-Injured without Litigation, and FP = Forensic Participants

Figure 14. *Magnitude of Error on the WCST by Group*

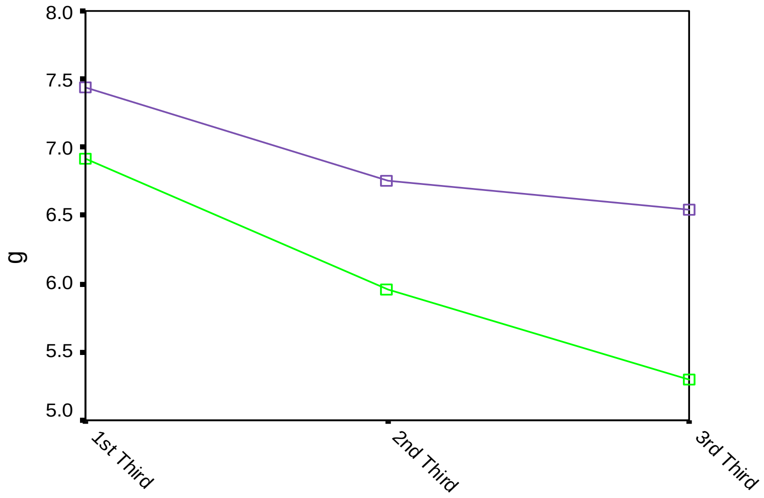


Hypothesis Three Inferential Statistics

The third hypothesis states that the rate of change in performance (rate of decay) on the WAIS-R NI subtests, the COWA, and the Finger oscillation for the malingering group should be smaller than for the head injured without litigation and control groups, yet similar to the rate of change in performance (rate of decay) for the head-injured with litigation and forensic group

To statistically test the difference in the rate of decay on the assessment instruments for the malingering group and honest responders, each assessment instrument was divided into three equal sections, with each section increasing in difficulty. The sections were then analyzed using a repeated measures MANOVA with linear polynomial contrasts. A group by section linear interaction test was performed on the WAIS-R NI subtests, the COWA, the Finger Oscillation test. Only one subtest was statistically significant at an alpha criterion of .05: The similarities subtest on the WAIS-R NI (See Figure 15) gave a statistically significant result ($F_{(1)} = 4.120$, $p = .045$, $\eta^2 = .032$, $1 - \beta = .522$). The mean slope for the malingerers group was $M = -1.14$; the honest responders group was $M = -.64$.

Figure 15. *Linear Interaction Contrast for WAIS-R NI Similarities Subtest*

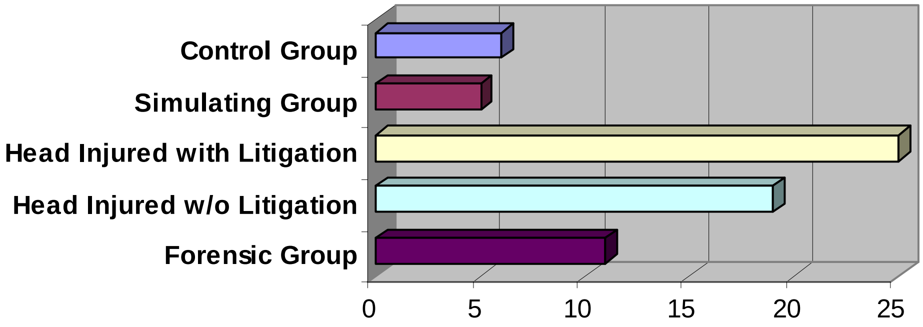


Hypothesis Four Inferential Statistics

The fourth hypothesis examined the validity of the 14-Item Performance Validity Index (VI) for the BCT. Bolter (1992) designed the index to aid in the detection of neuropsychological malingering.

To statistically test this hypothesis, a one-way ANOVA with multiple comparisons was computed on the total number of VI items endorsed on the BCT for each of the five groups. No significant between-group difference was found. The ANOVA analysis was followed by the Student-Newman-Keuls (SNK) post-hoc test statistic. The SNK failed to separate the items into subsets, finding them homogenous. Interestingly, the SP group actually had the lowest total errors (5), while the FP had a total of 11 errors (see Figure 16). It is noteworthy that the HL group had the most number of errors (25).

Figure 16. *Total Number of BCT IV Errors per Group*



Debriefing Questionnaire

The current study utilized a short answer and multiple choice questionnaire which was developed to examine select characteristics of the sample population. Only the HL and HN groups were excluded from completing the questionnaire, as their tests results were taken from a longer, full-day neuropsychological assessment and they were not volunteer participants in the current study.

Responses to the first two questions indicated that 43% of the participants understood most of the instructions, while 40% understood all of the instructions. 10% understood part of the instructions, while 6% did not really understand the instructions. However, when asked to restate the instructions, responses varied greatly. 27% indicated that the testing was to help complete a Ph.D. or for research purposes. 14% reinstated the actual directions for each of the measures, and only 14% were able to give a good synopsis of the initial instructions provided regarding the project. The remaining 47% gave mainly one sentence responses, such as “to help people with brain damage,” or “to do the best I could.”

When asked how hard they tried to follow the instructions, 6% indicated they tried moderately hard (about ½ of the time), 20% indicted they tried significantly hard (about ¾ of the time), and the vast majority (70%) indicated that they tried very hard on all tests. Level of brain knowledge was good for 39% of the participants, and adequate for 50% of the participants, but poor for 7% of the participants. The remaining 4% did not answer the question.

The CP and FP groups were asked if they felt they provided their best performance on each test. 36% indicated that they were somewhat successful, 29% were moderately successful, and 32% were very successful. Only one participant indicated that they were not successful at all, and the same individual was unable to restate the initial instructions. The SP group was asked which strategies they used in attempting to fake. Common answers were 1) “tried to miss a category” (RMT), 2) “slowed down giving words” (COWA), 3) “did not go as fast as I could” (Finger Oscillation), 4) “Tried to miss every couple questions”…”tried to miss one every now and then” (WAIS-R NI), 5) “took my time answering questions”…”tried to put answers now and then that I wouldn’t normally have put” (BCT), and 6) “missed one on purpose if I got the one before”…”didn’t spend a lot of time trying to figure it out like I would have if I wasn’t faking” (WCST).” Interestingly, for the WAIS-R NI subtests, RMT, COWA, and Finger oscillation, most simulators utilized strategies which would both decrease their performance below intact levels of functioning, and more likely result in a performance curve reflective of random responding. However, on both the BCT and WCST, the majority of the simulators utilized the strategy of extended latency responses, which would theoretically have no effect on the quality of their response. None of the simulators reported a strategy which would mimic the concept of magnitude of error.

CHAPTER IV

DISCUSSION

Few assessment instruments are available which provide accurate and consistent clinical information on malingered performances during neurocognitive evaluations. Those instruments that currently provide adequate cut-off scores typically assess the domain of attention and memory, such as the Test of Memory Malingering (TOMM; Tom Tombaugh, 1996), or the Word Memory Test (WMT; Paul Green, 2001). Tests like the TOMM and WMT are easily performed at a high level of accuracy by well-motivated individuals, even those with severe cognitive impairments, brain injury, or neurological diseases. Interestingly, individuals with mild head injuries tend to score lower on malingering, or “effort” measures than those with moderate to severe brain injury, leading one to question the level of effort or motivation expended during testing (Green, Iverson, & Allen, 1999). Likewise, Green, Rohling, Lees-Haley, & Allen (2001) suggest as well that individuals with mild head injuries perform far worse on “effort” measures than genuine patients with severe brain injuries or brain tumors. Aside from the few malingering measures available, psychologists and neuropsychologists who evaluate clients and patients typically rely on conflicting and/or inconsistent case histories, both between- and across-test discrepancies, base rates of true disease states compared to invalid or exaggerated test data, discrepancies between subjective complaints, test performance/results and behavioral observations, implausible symptoms such as remote memory loss, and comparisons across testing sessions to assist in malingering detection. Clinical judgment has also been used, however, as the current study has indicated, clinical judgment tends to be less than adequate, particularly when only a protocol is reviewed without concurrent client interview.

Further empirical studies are needed to clarify areas of uncertainty, such as base rates for malingering behavior, discrepancies between subjective complaints and actual impairment, and the development of “hold” tests in the face of malingering behavior that will strengthen diagnostic certainty. Also, the best methods for applying and clinically interpreting the testing results are still evolving, particularly given continued development of neurocognitive profiles for both neurological and neuropsychiatric disorders. Rogers (1988b) discussed malingering in terms of adaptive behavior (as opposed to the notion of malingering as a trait). In this vein, malingering is not a disorder, but rather a behavior, therefore it is not diagnosed, as with clinical conditions, but rather it is detected. There is the argument, however, that malingering may manifest in specific personality disorders, most notably antisocial personality disorder, and is therefore more a diagnostic criteria of the disorder rather than a resulting behavior.

The purpose of the current research was to obtain guidelines on six neuropsychological instruments which will allow for better identification, or detection, of malingering. Five population samples were selected: control participants (CP), simulating participants (SP), head injured patients with litigation (HL), head injured patients without litigation (HN), and forensic parole participants (FP). The current sample not withstanding, using simulators to represent known malingerers raises concerns regarding external validity. However, because retention of actual malingerers is so difficult, studies typically utilize simulators to validate instruments.

The current study sought to examine the viability of individuals to meet the criteria of generally agreed upon definitions of malingering. As discussed previously, using both Resnick=s and Roger’s concept of malingering and simulation, this research did not support their premises. On the contrary, the majority of simulators in the current study were not effective in eluding identification as a malingerer. The major reason for their easily identifiable results was their severely impaired performance, which would not be the expectation had they suffered a head injury and had no ulterior motives to malinger. In some cases, the severity of the malingered impairment was so great as to be blatantly unbelievable. The simulating group reported that their most frequently utilized strategy was to randomly respond, missing every few items. Interestingly, this is a common indicator utilized by neuropsychologists in the very detection of malingering. As mentioned in the results section, a surprising finding was the strategy of slowed verbal latency, or slowed responding. A number of the simulators reported that they utilized this strategy in an attempt to appear dysfunctional or impaired, without regard to the fact that they were still providing mostly accurate answers.

Incentives have been utilized in numerous research projects as an enticement for malingering. Rogers et al. (1993) indicated that without efforts to measure incentives and their effects upon the participants, as well as the participants motivation to follow instructions and consistently attempt to malinger believable deficits, the results from any research has limited generalizability to real-world malingerers. Based on the premise that participants asked to malinger should be provided an incentive for success, an incentive was provided in the current study. The incentive did not appear to have an significant effect upon the involvement of the CP and SP groups, but did appear to influence the involvement of FP participants. Both CP and SP groups appeared interested in the testing process, and concerned with their performance. In contrast, word of mouth that a monetary incentive was available for doing a few tests greatly increased the number of FP individuals who signed up for participation. The FP group also paralleled the CP and SP groups in regards to displaying interest in their performance, however, with the difference that they appeared less concerned about their failures, and more interested in their successes.

As indicated by Rogers (1988b) the verification of a participant’s motivation to follow instructions and feign a deficit is an important aspect in understanding malingering behavior. The method utilized in the current study to verify the motivation of the SP group was a debriefing questionnaire. A number of interesting pieces of information were gleaned from this questionnaire. First, most simulators did not feel confident that they had successfully eluded the examiner, even when they indicated a good knowledge of the brain. Second, the strategies they employed included methods by which clinicians attempt to identify suboptimal performance (e.g., random responding), as well as methods that would have no bearing on the quality of their answers, but would further provide incongruent information regarding behavior versus test results (e.g., slowed verbal latencies). In general, the simulators in the current study did not appear “sophisticated enough” to successfully malinger. This point lends itself to the concept that experienced malingerers, those individuals who have been in the system and have learned more efficient ways of successfully feigning, are not likely to be the “typical citizen” who becomes involved in an accident which leads to a mild head injury or post-concussive syndrome. This would beg the question as to whether malingering detection in unsophisticated patients is even necessary. A clinician’s job is not only to determine what information is relevant to the diagnostic question but also to determine the relative usefulness of the information. In support of that supposition, in the current study, participants who were instructed to malinger consistently performed more poorly (displayed considerably more impairment) than patients with acquired brain damage and forensic participants alike, with statistically significant effect size differences which often exceeded two standard deviations.

Cutoff Scores and Hit Rates

Cutoff score ranges for suspecting malingering have been provided for the instruments used in the present study. Table 18 (p. 130) displays a range of cutting scores, as well as the sensitivity, specificity, PPV, NPV, hit rates, and total number of false negatives, true positives, false positives, and true negatives per cut score per assessment instrument**.** Tables 12 through 17 (pp. 124) indicate cutoff scores by assessment instruments which are believed to be optimal based on the relationship between hit rates, PPV, and NPV. Using data from the present study for the purpose of assisting in malingering detection should be done both quantitatively and qualitatively. Current research suggests that a single cutting score may not accurately identify malingerers. Clinicians should review the tables and figures provided and use discretion in their decision-making, taking care to avoid committing false-positive errors.

As mentioned previously, the subtests on the WAIS-R NI (Wechsler Adult Intelligence Scale-Revised® Neuropsychological Instrument; WAIS-R® NI; Psychological Corporation, San Antonio, TX, www.psychcorp.com) showed promise in being able to detect malingered performances, as also did the Finger Oscillation test (HRNB; Reitan & Wolfson, 1993). The current study revealed that simulators appear to over exaggerate motor deficits, resulting in very poor scores averaging around 35 mean number of taps, even as the forensic population performed the best, obtaining the highest number of mean taps. Although this difference appears extreme in comparison to the remaining four groups, it parallels preliminary findings in a study by Larrabee (2002). He proposed a combined right plus left hand raw finger tapping score of less than 63 as one indicator of malingering. A score of less than 63 total taps bilaterally identified 10 of 25 (40%) of patients with definite malingered neurocognitive dysfunction, and correctly identified 29 of 31 participants (93.5%) with moderate to severe closed head injury. Larrabee's findings and our research are nearly identical, indicating a possible cut score of < 64 mean taps bilaterally for use in malingering detection.

Given prior research assessing the sensitivity and specificity of the RMT (Rey-15 Item Memory Test; RMT; Andre Rey, 1964), and WCST (Wisconsin Card Sorting Test^{TM}; WCST^{TM};

Psychological Assessment Resources, Lutz, FL, www.parinc.com), our findings, although interesting, were not altogether surprising. Using a cutoff score of 9, the current study obtained a hit rate of only 37.3% on the RMT, with a sensitivity of 84%, but a specificity of only 64% and a 56.3% false-positive rate. In comparison, Greiffenstein et al. (1994) obtained 71% correct classification for both the brain-injury and malingering groups, however, they did not specify the number of false-positives identified. Our research was also less effective in distinguishing malingerers than both Simon (1994), who correctly classified 85.7% of his simulating participants, with a 14.3% false positive rate for controls when using 9 as the cut score, and Shretlan et al. (1991), who misclassified 21.6% of both their patient and simulating population. When a cutoff score of 8 was utilized in the current study, our sensitivity ratio approached 95%, but sensitivity decreased considerably to 26%, with a 59% false-positive rate. When Bernard et al (1990) reduced the cutting score to 8, they correctly classified 100% of their population. However, we also examined the efficacy of total number of groups recalled to determine if cutoff scores could accurately predict malingering behavior. Using total groups recalled, the SP group displayed the fewest number of categories recalled, while interestingly, both head injured groups displayed the most categories recalled, thought not significantly so when compared to the CP and FP groups. Based on the current findings, the RMT continues to be an unreliable method for malingering determination.

Likewise, findings with the WCST revealed poor specificity. These results are consistent with recent research undertaken by Bernard, McGrath, and Houston (1996), and Suhr and Boyer (1999). Bernard et al. (1996) utilized number of categories and perseverative errors as research variables, obtaining a specificity ratio of 54.3%. Suhr and Boyer (1999) obtained a specificity ratio of 38.3% utilizing number of categories and failure to maintain set as research variables.

Magnitude of Error

Examination of the magnitude of error theory in the present study revealed promising results, both for the concept and for its application to the WAIS-R NI Vocabulary, Similarities, and Information subtests, and the WCST. Utilizing these instruments, the SP group consistently performed poorer than the CP and HN groups, paralleled only by the FP group. Interestingly, the HL group was similar to the honest responders, and in some cases actually performed better.

A second interesting finding was the poor utility of the magnitude of error theory when using the BCT (Booklet Category Test; HRNB; Reitan & Wolfson, 1993). Utilizing this concept, the BCT was better able to identify actual head-injured populations in contrast to malingering behavior. Both the HL and HN groups performed consistently worse, rivaled only by the FP group. Subtest differences did emerge, however, with some interesting results. On subtest one, which consists of a very easy to grasp strategy and is the shortest subtest, neither the CP nor FP group made errors. Both the HL and HN group made a similar number of errors, followed by the SP group. Similarly, on subtest two, which is also relatively easy to grasp as compared to subtests 3 through 6, the SP group made a significant number of errors, more so than the CP, HL and HN groups combined. However, across subtests 3 through 6, the number of errors for the SP group gradually declined, while the performance for the CP, HL and HN groups was more variable, and the FP group displayed the greatest difficulty overall. Finally, on subtest three, which is generally considered one of the most, if not the most, difficult of the subtests, the HL, HN and FP groups displayed the poorest performances (see Figure 23).

Performance Curve

Careful examination of a patient’s performance curve, or rate of decay, allows the clinician to determine if a particular performance obtained from a patient is rare or unusual as compared to known groups. Generally, one should expect to see perfect performance, followed by a decrement in performance, transitioning to 50% performance accuracy. This gives the curve a characteristic shape for compliant performance without respect to the ability of the test taker. Using performance curve characteristics as a means of examining the potential for response invalidity has received endorsement for cognitive tests (Gudjonsson & Shackleton, 1986; McKinzey, Podd, Krehbeil, & Raven, 1999; Rogers, et al., 1993). Unfortunately, the current study was not effective in advancing the research on performance curve characteristics. Out of the three tests examined in the current study, only one displayed moderate promise in detection of altered performance curves, the WAIS-R NI Similarities subtest.

Strengths of the Current Study

Well developed research studies attempt to minimize the effects of extraneous variables, along with minimizing the effects of random variables. In the present study, unique group membership was equivocally necessary in order to provide meaningful data. Therefore, several factors were considered when assigning group membership. First, in order to address effort issues regarding possible future compensation, two head-injured groups were included, one which was involved in litigation, and one which was not. Second, most previous research utilized college students only as simulating participants, which inherently reduces age ranges, education levels, and generational factors, such as having started a family or establishing a career. In the current study, community control subjects who were asked to malinger were compared to community control subjects who were asked to provide their best performance, which allowed for a more accurate representation of age, education, and experience parallel with the general population. Third, it is also advantageous to include forensic patients within a study as a further standardization vehicle, and to allow cutting scores to be examined for that portion of the population which is most likely to malinger. The current study included parolees who had committed a felony offense, had recently been released from prison for that offense, and who had presented with negative results on their two most recent urine analysis examinations, including one on the day of testing.

We attempted to control for random error in several ways. When the study was started, only those instruments which were commonly used at that time were included. The majority of data was collected within a 2-3 year period, with no data collection exceeding 3.5 years. Standardization was maintained across testing sites between examiners by using an agreed-upon testing order, and the same administration instructions. Additionally, with the exception of a number of the head-injured sample, the CP, SP, and FP populations were all administered the assessment instruments by the major researcher of this project.

Researchers have indicated that types of coaching may have an influence on the success of malingers (Berry et al., 1994). Aside from the instructions to malinger a believable injury, no coaching was provided to the SP group. As a point of fact, the examiner was blind as to group membership, as the control and simulating groups were randomly assigned via a sealed envelop and therefore the examiner was not able to provide additional verbal coaching and/or instructions during the testing session. Written instructions for these two groups were also provided to reduce administration error. Specific instructions for the control group requested participants to respond in an honest nature, while instructions for the simulating group asked participants to feign a head injury from an automobile accident (see Appendix C).

Limitations of the Current Study

Although statistical power was not an issue in the present study, it is important to note that the data obtained from the clinical groups is based on smaller sample sizes. Replication of these base rates on medium to large samples is recommended.

A second limiting factor consistent in most malingering research is the inability to obtain a known group of malingerers. In the current study, a group of forensic participants, specifically, males who had been incarcerated for felony offenses and were recently released on parole were utilized as a presumed known group of malingerers. Testing results and behavioral observations for this particular group, however, revealed variability in effort as well as ability in completing the assessment instruments. Although as a group, the results for the forensic participants often approximated that of the simulators versus the presumed honest responders, the variable regarding intentional effort in test taking for this group remains unknown.

Third, demographic variables may have confounded some of the findings. For example, the forensic group consisted entirely of males, while the head-injured without litigation consisted of twenty-two males and only 3 females. Similarly, the majority of group members consisted of Caucasians, with the exception of the forensic group, which was made up mainly of African-Americans. Educational level, though less of a difference than ethnicity, was still not as closely matched as would have been liked. The forensic group displayed the lowest total years of education, which likely adversely affected estimated IQ. This could be particularly informative, given that a number of researchers postulate that the construct of effort should be unrelated to age and IQ levels (P. Green, personal communication, April 11, 2003). Obtaining group memberships which more closely match each other across variables may produce different findings.

Finally, as discussed previously, choosing more recent and commonly-employed measures would better benefit clinicians in the field. There is an inherent difficulty with this ideal, however, as testing measures are frequently and sometimes almost concurrently being updated. This implicates the need for malingering research to continually repeat analyses with updated or new measures in order to provide validity to malingering detection.

Future Directions for Neuropsychological Malingering Research

According to Rogers, there are a number of largely ignored issues which need to be addressed when conducting malingering research: "(a) differentiating between malingerers and severely disturbed psychiatric patients with atypical presentations, (b) developing correction formulas for different styles of dissimulation, and (c) developing and validating standardized measures for gradations of dissimulation@ (1988b, p. 309). Correction formulas allow for further assessment of patients into classifications such as defensiveness and malingering (Rogers, 1988b). Correction scores also help in developing measures for gradations of dissimulation; such measures may allow for the detection of motivation towards dissimulation. As these needs remain important in malingering research, several newer, current issues will be discussed here.

Malingering Standards Across Cognitive Domains

As mentioned previously, most assessment instruments developed specifically to address issues of effort or malingering tend to measure attention or memory. It is noteworthy that memory can be one of the domains easily affected by brain injury or disease, however, there are other domains which would also be amenable to malingering standards. A multitude of tests are sensitive to effort issues. In Sweet's book, Forensic Psychology: Fundamentals and Practice (1999), he lists tests specifically designed for effort, as well as common neuropsychological measures sensitive to insufficient effort. Not only are the memory measures listed, but he also provides measures typically associated with visual-spatial, motor, sensory-perceptual, and problem-solving skills. Larrabee (2002) is also currently working on developing standard scores which would discriminate between persons identified with definite malingered neurocognitive dysfunction, and persons with traumatic brain injury. The assessment instruments he is utilizing in his research are Benton's Visual Form Discrimination, Combined (right plus left hand) Raw Finger Tapping (as discussed previously), Digit Span, Wisconsin Card Sorting Failure to Maintain Set, and the Lees-Haley Fake Bad Scale of the MMPI-2. He has recently submitted his findings for publication (G. Larrabee, personal communication, May 1, 2003).

In addition to attempting to malinger deficits on a specific domain, deception can also occur by reporting only one symptom falsely (e.g. back pain). In this context, some individuals may exaggerate various symptoms but not exaggerate cognitive difficulties. Also inherent in developing malingering standards on existing cognitive and psychological tests is the necessity to eliminate concerns regarding patient confusion. Utilizing below chance scores on forced choice tests, such as recognition trails on memory tests, assists in providing directions which allow little room for confusion. Future malingering research will need to remain sensitive to confounding variables, such as real or feigned confusion regarding assessment instructions, and symptom exaggeration versus malingering within cognitive domains.

Research and Base Rates

An important undertaking for neuropsychologists is to contribute to the collection of disorder base rates. Base rate indicates the prevalence of a sign or disorder within a specified population. However, any assessment instrument utilized clinically for malingering detection must yield a high sensitivity percentage (ability to identify malingering when it is present) while also yielding a high specificity percentage (which would reduce false-positives). Yet, sensitivity and specificity are not impacted by the base rate of the disorder in question. ”In contrast to sensitivity and specificity, PPV and NPV are affected by the base rate of the disorder in question,” (Labarge, McCaffrey, and Brown, 2003). Therefore, the authors speculate that the extensive use of neuropsychological instruments to assess and diagnose gives rise to the importance of base rates, particularly given the increasingly need to improve upon neuropsychologists' interpretation of diagnostic tests using techniques such as likelihood and odds ratios (Labarge et al., 2003).

Interestingly, Labarge et al's research indicated that neuropsychologists who participated in their study either "neglected or misused base rate information," including making inaccurate calculations of PPV when presented in a probability format (2003). Even when provided in a frequency format, 35.3% of the participants who correctly answered the PPV question incorrectly answered questions regarding base rate, sensitivity or specificity. Sackett, Richardson, Rosenberg & Haynes provide information regarding an internet site they are developing regarding medical base rate data (1997). To date, no one has undertaken a similar site for neuropsychology; however, a compilation of published symptom base rates for psychological, neuropsychological, and neurologic disorders is currently being prepared for publication (Labarge et al., 2003).

Coaching and Training

Other directions for malingering research would point to development of instruments which are relatively, if not entirely, resistant to training or coaching, even for individuals with some training in psychological assessment. In particular are attorneys, who may provide litigants with information and/or response styles of genuinely impaired individuals for particular interview or assessment protocols. In some cases, simply warning a client that tests will be given that address suboptimal or "malingering" performance may result in more sophisticating responding which may elude malingering indices.

Multiple Measures

Research indicates that clinicians might expect to have disagreement about effort in about 20% of cases using any two effort measures. Passing one effort measure does not imply that another effort measure would also be passed, as various measures differ in sensitivity and specificity. Another variable is the consistency of individuals across testing, and between testing evaluations. Good performance on "effort" tests does not necessarily rule out suboptimal performance on other portions of the exam. Therefore, when a patient performs poorly on an effort test, it suggests that he or she was not applying best effort at that particular point in time. A prudent clinician will then exercise caution when interpreting of the remainder of the patient's test results, exploring all possible explanations for the data. Multiple effort measures are needed to measure inconsistent effort.

Based on the current research findings, the WAIS-R NI Vocabulary, Information, and Similarities subtests and Finger Oscillation provided the strongest indications for use of cutting scores to assist in malingering detection. Interestingly, the WCST provided the strongest results when examining the magnitude of error strategy, while the WAIS-R NI subtests provided adequate results. In general, however, a good clinician does not analyze neurocognitive data in isolation, and diagnostic formulations need to make sense in the context of the nature of the injury/disease, clinical course, and adaptive ability level of the patient when attempting to “detect’ malingering behavior.

APPENDIX A

Informed Consent - Honest Responder

The psychology department at the University of North Texas is conducting research designed to determine the effects of pretending to be brain damaged on a group of psychological tests. I have been asked to take part as a person responding honestly to the examiner. **I will not inform any individuals involved with this study as to my role during the test taking phase: as an honest responder.**

If I agree to take part in this research, I will be asked to complete various neuropsychological tests that involve listening, looking, and answering brief questions in accordance with the instructions provided by the researcher. I will also be asked to provide brief information regarding my age, sex, socio-economic status, schooling (e.g., number of years, degree obtained), marital status, current employment, and physical health (e.g., past and current major illnesses or injuries). The entire study will take approximately two hours of my time.

There will be no direct benefit to me from these testing procedures, but the information gained may in the future help other people with head injuries and other neurological conditions. The study has been explained to me, and I have received and understand the instructions provided by the researcher and my role in this study.

All research records will be confidential. My records will be identified by a number known only to the primary investigator and her associates. My name will not be used in any paper or publications that may arise from this research.

Participation in research is entirely voluntary. I may refuse to participate or withdraw at any time without jeopardy. If I choose to withdraw, I will not inform others of the study. I will be paid $10.00 for my completed participation in this study. If I choose to withdraw, I will not be paid. Additionally, a monetary lottery will be held upon completion of data collection. One individual will be drawn from a pool, and rewarded $50.00 for their participation in this study. As a member of the honest responding group, my code number will be added to this lottery pool.

I have received a copy of this document and "The Experimental Subject's Bill of Rights". I agree to participate:

______________________________	______________________________
Participant's Signature/Date	Witness' Signature/Date

APPENDIX B

Informed Consent – Simulator

The psychology department at the University of North Texas is conducting research designed to determine the effects of pretending to be brain damaged on a group of psychological tests. I have been asked to take part as a person pretending to have a head injury from an automobile accident. **I will not inform any individuals involved with this study as to my role during the test taking phase: as a brain injured individual.**

If I agree to take part in this research, I will be asked to complete various neuropsychological tests that involve listening, looking, and answering brief questions in accordance with the instructions provided by the researcher. I will also be asked to provide brief information regarding my age, sex, socio-economic status, schooling (e.g., number of years, degree obtained), marital status, current employment, and physical health (e.g., past and current major illnesses or injuries). The entire study will take approximately two hours of my time.

There will be no direct benefit to me from these testing procedures, but the information gained may in the future help other people with head injuries and other neurological conditions. The study has been explained to me, and I have received and understand the instructions provided by the researcher and my role in this study.

All research records will be confidential. My records will be identified by a number known only to the primary investigator and her associates. My name will not be used in any paper or publications that may arise from this research.

Participation in research is entirely voluntary. I may refuse to participate or withdraw at any time without jeopardy. If I choose to withdraw, I will not inform others of the study. I will be paid $10.00 for my completed participation in this study. If I choose to withdraw, I will not be paid. Additionally, a monetary lottery will be held upon completion of data collection. One individual will be drawn from a pool, and rewarded $50.00 for their participation in this study. As a member of the simulation group, my code number will be added to this lottery pool.

I have received a copy of this document and "The Experimental Subject's Bill of Rights". I agree to participate:

______________________________	______________________________
Participant's Signature/Date	Witness' Signature/Date

APPENDIX C

Malingering Instructions for Simulation Group

You are now going to take six neuropsychological tests, which measure various cognitive functions, as part of a research project. Your part in this project is to take these tests as you think a person with a brain injury would take them. Pretend you have suffered a head injury six months ago in an automobile accident caused by someone else. You were unconscious for a period of time, and were hospitalized overnight. Because the accident was caused by someone else, you are involved in litigation (a lawsuit) to pursue financial compensation for your injuries, from the person responsible and the insurance companies involved. You are now a brain-damaged individual. Imagine that your everyday functioning, such as school or work activities, has declined significantly since your accident. Due to your currently poor work habits, your earning power has been substantially reduced, and therefore, you deserve all the money that the court will allow for your case. Your test results will help determine how large your settlement will be. Therefore, fake the most severe, yet realistic disabilities that you can to try to deceive me into thinking you have a brain injury. However, don't go overboard and overplay your role or be too obvious, or I'll know you are trying to deceive me. Feel free to alter your performance in any way you see fit to produce believable, realistic results. It is important that you do not tell the testing examiner what these instructions ask you to do.

APPENDIX D

Informed Consent - Honest Responder - Parolee

The psychology department at the University of North Texas is conducting research designed to determine the effects of pretending to be brain damaged on a group of psychological tests. I have been asked to take part as a person responding honestly to the examiner.

If I agree to take part in this research, I will be asked to complete various neuropsychological tests that involve listening, looking, and answering brief questions in accordance with the instructions provided by the researcher. I will also be asked to provide brief information regarding my age, sex, socio-economic status, schooling (e.g., number of years, degree obtained), marital status, current employment, and physical health (e.g., past and current major illnesses or injuries). The entire study will take approximately two hours of my time.

There will be no direct benefit to me from these testing procedures, but the information gained may in the future help other people with head injuries and other neurological conditions. The study has been explained to me, and I have received and understand the instructions provided by the researcher and my role in this study.

Participation in research is entirely voluntary. I may refuse to participate or withdraw at any time without jeopardy. If I choose to withdraw, I will not inform others of the study. I will be paid $10.00 for my completed participation in this study. If I choose to withdraw, I will not be paid.

All research records will be confidential. My records will be identified by a number known only to the primary investigator and her associates. My name will not be used in any paper or publications that may arise from this research. The information obtained will not be provided to the Parole Board, or my probation officer, and will not be used against me in any way.

For the purposes of this study only, I voluntarily answer the following two questions, both of which fall under the confidential nature of this research, as discussed above.

1) Please describe any past (not current) substance abuse (for both illegal and legal substances): ____
____
____
____

2) Please indicate how much and how long for each substance listed in #1 above:
____
____
____
____

I have received a copy of this document and "The Experimental Subject's Bill of Rights". I agree to participate:

____ ______________________________________

Participant's Signature/Date Witness' Signature/Date

APPENDIX E

Post-Research Questionnaire - Simulator

Participant #: ________________ Date: ___________________ Technician: ____________

1) How well do you feel you understood the instructions provided to you?
   a) Did not really understand the instructions
   b) Understood part of the instructions
   c) Understood most of the instructions
   d) Understood all of the instructions

2) Would you please restate the instructions you received at the beginning of this testing?

____

____

____

____

____

____

3) How hard did you try to follow these instructions?
   a) Did not try at all
   b) Tried somewhat (about 1/4 the time)
   c) Tried moderately (about ½ the time)
   d) Tried significantly (about 3/4 the time)
   e) Tried very hard on all tests

4) Which tests did you attempt to fake? (Circle all that apply):
   1) REY
   2) COWA
   3) Tapping
   4) BCT
   5) WCST
   6) WAIS-R NI

5) How successful do you think you were in faking results of a believable brain injury and/or brain damage?
   a) Not at all successful
   b) Somewhat successful
   c) Moderately successful
   d) Significantly successful
   e) Very successful

6) If you feel you were successful in faking a head injury, what helped you to fake? (Circle all that apply).
   a) Knowledge of the brain
   b) Have known people with brain injury or brain damage
   c) Able to follow instructions well.
   d) I am a quick learner.

e) Any additional reasons that helped you fake a brain injury?

____

____

7) At what level do you think your knowledge of brain functioning is?

a) Highly specialized
b) Good
c) Adequate
d) Poor
e) Other:

____

____

____

8) Do you believe you were successful in keeping the examiner from discovering you were malingering?

a) Not at all successful
b) Somewhat successful
c) Moderately successful
d) Very successful

9) For each test you took, please provide those strategies or methods you utilized in your attempt to malinger a head injury with believable deficits:

Test #1 (RMT):____
____
Test #2 (COWA): ____
____
Test #3 (Tapping): ____
____
Test #4 (BCT): ____
____
Test #5 (WCST): ____
____
Test #6 (WAIS-R NI): ____
____

10) If you do not feel you were able to fake well, what hampered you?

a) I am too honest.
b) I didn’t understand the instructions.
c) The tests were too easy.
d) The tests were too hard.
e) Other reasons: ____

____

APPENDIX F

Post-Research Questionnaire - Honest Responder

Participant #: ________________ Date: _________________ Technician:______________

1) How well do you feel you understood the instructions provided to you?
   - a) Did not really understand the instructions
   - b) Understood part of the instructions
   - c) Understood most of the instructions
   - d) Understood all of the instructions

2) Would you please restate the instructions you received during this testing?

   ____

   ____

   ____

   ____

   ____

   ____

3) How hard did you try to follow these instructions?
   - a) Did not try at all
   - b) Tried somewhat (about 1/4 the time)
   - c) Tried moderately (about ½ the time)
   - d) Tried significantly (about 3/4 the time)
   - e) Tried very hard on all tests

4) At what level do you think your knowledge of brain functioning is?
   - a) Highly specialized
   - b) Good
   - c) Adequate
   - d) Poor
   - e) Other:

   ____

   ____

   ____

5) For each test you took, please provide those strategies or methods you utilized in your attempt to provide your best performance:

Test #1 (RMT): ______________________________

Test #2 (COWA): ______________________________

Test #3 (Tapping): ______________________________

Test #4 (BCT): ______________________________

Test #5 (WCST): ______________________________

Test #6 (WAIS-R NI): ______________________________

6) Do you believe you were successful in providing your best performance on each test?

a)	Not at all successful	c)	Moderately successful
b)	Somewhat successful	d)	Very successful

7) If you do not feel you were able to provide your best performance, what hampered you?
a) I didn’t understand the instructions.
b) The tests were too easy.
c) The tests were too hard.
d) Other reasons:______________________________

Table 12: *Efficacy of the WAIS-R NI Subtests*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Vocabulary Subtest	33	40.5%	1.000	0.013	40.0%	100.0%
T-Score	36	50.8%	0.980	0.197	44.5%	93.8%
	40	62.7%	0.860	0.474	51.8%	83.7%
	43	69.8%	0.68	0.711	60.7%	77.1%
	46	70.6%	0.5	0.842	67.6%	71.9%
Information Subtest	33	42.1%	0.980	0.053	40.5%	80.0%
T-Score	36	52.4%	0.960	0.237	45.3%	90.0%
	40	60.3%	0.92	0.395	50.0%	88.2%
	43	66.7%	0.92	0.5	54.8%	90.5%
	46	69.8%	0.76	0.658	59.4%	80.6%
Similarities Subtest	33	42.3%	1.000	0.013	41.9%	100.0%
T-Score	36	48.5%	1.000	0.118	44.6%	100.0%
	40	50.8%	0.981	0.171	45.7%	92.9%
	43	59.2%	0.981	0.316	50.5%	96.0%
	46	63.8%	0.944	0.421	53.7%	91.4%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 13: *Efficacy of the RMT*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Rey Total Items	7	38.1%	0.960	0.000	38.7%	0.0%
	8	38.9%	0.940	0.026	38.8%	40.0%
	9	37.3%	0.840	0.066	37.2%	38.5%
	10	38.9%	0.820	0.105	37.6%	47.1%
	11	45.2%	0.780	0.237	40.2%	62.1%
	12	57.1%	0.72	0.474	47.4%	72.0%
	13	55.6%	0.62	0.513	45.6%	67.2%
Rey Total Groups	2	38.1%	0.920	0.026	38.3%	33.3%
	3	50.8%	0.780	0.329	43.3%	69.4%
	4	60.3%	0.560	0.632	50.0%	68.6%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 14: *Efficacy of the COWA*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
COWA Total Words	25	49.2%	0.780	0.303	42.4%	67.6%
	26	49.2%	0.740	0.329	42.0%	65.8%
	27	51.6%	0.7	0.395	43.2%	66.7%
	28	54.8%	0.66	0.473	45.2%	67.9%
	29	55.6%	0.640	0.500	45.7%	67.9%
	30	57.9%	0.640	0.539	47.8%	69.5%
	31	55.6%	0.520	0.579	44.8%	64.7%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 15: *Efficacy of Finger Oscillation*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Total Mean Taps Dominant Hand	33	44.4%	0.9	0.145	40.9%	68.8%
	34	42.9%	0.88	0.132	40.0%	62.6%
	35	43.7%	0.88	0.145	40.4%	64.7%
	36	44.4%	0.84	0.184	40.4%	63.6%
	37	46.0%	0.8	0.237	40.8%	64.3%
	38	48.4%	0.76	0.303	41.8%	65.7%
	39	50.8%	0.76	0.342	43.2%	68.4%
	40	51.6%	0.72	0.382	43.4%	67.4%
	41	52.4%	0.68	0.421	43.6%	66.7%
	42	54.8%	0.560	0.539	44.4%	65.1%
	44	55.6%	0.520	0.579	44.8%	64.7%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 16: *Efficacy of the BCT*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Total Number Errors	58	50.0%	0.740	0.342	42.5%	66.7%
	59	50.0%	0.720	0.355	42.4%	35.9%
	60	48.4%	0.680	0.355	41.0%	62.8%
	61	47.6%	0.660	0.355	40.0%	61.4%
	62	47.6%	0.640	0.368	40.0%	60.9%
	64	47.6%	0.640	0.382	39.7%	60.4%
	66	50.8%	0.620	0.434	41.9%	63.5%
	67	50.8%	0.600	0.447	41.7%	63.0%
	68	51.6%	0.600	0.461	42.3%	63.6%
	69	50.8%	0.580	0.461	41.4%	62.5%
	70	51.6%	0.580	0.474	42.0%	63.2%
	71	51.6%	0.560	0.487	41.8%	62.7%
	73	50.8%	0.540	0.487	40.9%	61.7%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 17: *Efficacy of the WCST*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Total Number Errors	50	44.4%	0.120	0.658	18.8%	53.2%
	52	46.0%	0.100	0.697	17.9%	54.1%
	56	46.8%	0.080	0.724	16.0%	54.5%
	55	47.6%	0.080	0.737	16.7%	54.9%
	56	49.2%	0.080	0.763	18.2%	55.8%
	59	50.0%	0.080	0.776	19.0%	56.2%
	60	50.8%	0.080	0.789	20.0%	56.6%
	62	51.6%	0.080	0.803	21.1%	57.0%
	63	50.8%	0.060	0.803	16.7%	56.5%
	65	52.4%	0.060	0.829	18.8%	57.3%
	66	53.2%	0.060	0.842	20.0%	57.7%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 18: *Cutting Scores per Assessment Instrument*

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
WAIS-R NI Vocabulary T-Score	33	1.000	0.013	40.0%	100.0%	40.5%	0	50	75	1	126
	36	0.980	0.197	44.5%	93.8%	50.8%	1	49	61	15	126
	40	0.860	0.474	51.8%	83.7%	62.7%	7	43	40	36	126
	43	0.68	0.711	60.7%	77.1%	69.8%	16	34	22	54	126
	46	0.5	0.842	67.6%	71.9%	70.6%	25	25	12	64	126
	50	0.3	0.934	75.0%	67.0%	68.3%	35	15	5	71	126
WAIS-R NI Information T-Score	29	0.980	0.013	39.5%	50.0%	39.7%	1	49	75	1	126
	33	0.980	0.053	40.5%	80.0%	42.1%	1	49	72	4	126
	36	0.960	0.237	45.3%	90.0%	52.4%	2	48	58	18	126
	40	0.92	0.395	50.0%	88.2%	60.3%	4	46	46	30	126
	43	0.92	0.5	54.8%	90.5%	66.7%	4	46	38	38	126
	46	0.76	0.658	59.4%	80.6%	69.8%	12	38	26	50	126
	50	0.66	0.803	68.8%	78.2%	74.6%	17	33	15	61	126
WAIS-R NI Similarities T-Score	33	1.000	0.013	41.9%	100.0%	42.3%	0	54	75	1	130
	36	1.000	0.118	44.6%	100.0%	48.5%	0	54	67	9	130
	40	0.981	0.171	45.7%	92.9%	50.8%	1	53	63	13	130
	43	0.981	0.316	50.5%	96.0%	59.2%	1	53	52	24	130
	46	0.944	0.421	53.7%	91.4%	63.8%	3	51	44	32	130
	50	0.925	0.487	56.2%	90.2%	66.9%	4	50	39	37	130
RMT - Total Items	7	0.960	0.000	38.7%	0.0%	38.1%	2	48	76	0	126
	8	0.940	0.026	38.8%	40.0%	38.9%	3	47	74	2	126
	9	0.840	0.066	37.2%	38.5%	37.3%	8	42	71	5	126
	10	0.820	0.105	37.6%	47.1%	38.9%	9	41	68	8	126
	11	0.780	0.237	40.2%	62.1%	45.2%	11	39	58	18	126

(*table continues*)

Table 18 (*continued*):

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
RMT - Total Items	12	0.72	0.474	47.4%	72.0%	57.1%	14	36	40	36	126
(continued)	13	0.62	0.513	45.6%	67.2%	55.6%	19	31	37	39	126
	14	0.56	0.632	50.0%	68.6%	60.3%	22	28	28	48	126
RMT - Total Groups	2	0.920	0.026	38.3%	33.3%	38.1%	4	46	74	2	126
	3	0.780	0.329	43.3%	69.4%	50.8%	11	39	51	25	126
	4	0.560	0.632	50.0%	68.6%	60.3%	22	28	28	48	126
COWA - Total Raw Words	16-20	0.860	0.132	39.4%	58.8%	42.1%	7	43	66	10	126
Frequency Ranges	21-22	0.820	0.184	39.8%	60.9%	43.7%	9	41	62	14	126
	23-24	0.780	0.276	41.5%	65.6%	47.6%	11	39	55	21	126
23-24 Borderline	25-26	0.740	0.316	41.6%	64.9%	48.4%	13	37	52	24	126
25-30 Low Average	27-28	0.660	0.461	44.6%	67.3%	54.0%	17	33	41	35	126
31-44 Average	29-30	0.640	0.526	47.1%	69.0%	57.1%	18	32	36	40	126
45-52 High Average	31-32	0.500	0.618	46.3%	65.3%	57.1%	25	25	29	47	126
	33-34	0.460	0.658	46.9%	64.9%	57.9%	27	23	26	50	126
	35-36	0.360	0.776	51.4%	64.8%	61.1%	32	18	17	59	126
	37-38	0.280	0.803	48.3%	62.9%	59.5%	36	14	15	61	126
COWA - Total Raw Words	20	0.860	0.132	39.4%	58.8%	42.1%	7	43	66	10	126
	21	0.840	0.158	39.6%	60.0%	42.9%	8	42	64	12	126
	22	0.820	0.184	39.8%	60.9%	43.7%	9	41	62	14	126
	23	0.800	0.224	40.4%	63.0%	45.2%	10	40	59	17	126
	24	0.780	0.276	41.5%	65.6%	47.6%	11	39	55	21	126
	25	0.780	0.303	42.4%	67.6%	49.2%	11	39	53	23	126
	26	0.740	0.329	42.0%	65.8%	49.2%	13	37	51	25	126
	27	0.7	0.395	43.2%	66.7%	51.6%	15	35	46	30	126
	28	0.66	0.473	45.2%	67.9%	54.8%	17	33	40	36	126
	29	0.640	0.500	45.7%	67.9%	55.6%	18	32	38	38	126
	30	0.640	0.539	47.8%	69.5%	57.9%	18	32	35	41	126

(*table continues*)

Table 18 (*continued*):

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
COWA - Total Raw Words (continued)	31	0.520	0.579	44.8%	64.7%	55.6%	24	26	32	44	126
	32	0.500	0.632	47.2%	65.8%	57.9%	25	25	28	48	126
	33	0.480	0.645	47.1%	65.3%	57.9%	26	24	27	49	126
	34	0.460	0.671	47.9%	65.4%	58.7%	27	23	25	51	126
	35	0.380	0.724	47.5%	64.0%	58.7%	31	19	21	55	126
Finger Oscillation Dominant Mean Raw	30	0.92	0.132	41.1%	71.4%	44.4%	4	46	66	10	126
	31	0.92	0.132	41.1%	71.4%	44.4%	4	46	66	10	126
	32	0.9	0.118	40.2%	64.3%	42.9%	5	45	67	9	126
	33	0.9	0.145	40.9%	68.8%	44.4%	5	45	65	11	126
	34	0.88	0.132	40.0%	62.5%	42.9%	6	44	66	10	126
	35	0.88	0.145	40.4%	64.7%	43.7%	6	44	65	11	126
	36	0.84	0.184	40.4%	63.6%	44.4%	8	42	62	14	126
	37	0.8	0.237	40.8%	64.3%	46.0%	10	40	58	18	126
	38	0.76	0.303	41.8%	65.7%	48.4%	12	38	53	23	126
	39	0.76	0.342	43.2%	68.4%	50.8%	12	38	50	26	126
	40	0.72	0.382	43.4%	67.4%	51.6%	14	36	47	29	126
	41	0.68	0.421	43.6%	66.7%	52.4%	16	34	44	32	126
	42	0.560	0.539	44.4%	65.1%	54.8%	22	28	35	41	126
	44	0.520	0.579	44.8%	64.7%	55.6%	24	26	32	44	126
	45	0.38	0.645	41.3%	61.3%	54.0%	31	19	27	49	126
	46	0.32	0.684	40.0%	60.5%	54.0%	34	16	24	52	126
	47	0.28	0.75	42.4%	61.3%	56.3%	36	14	19	57	126
Finger Oscillation Dominant Mean Raw (excluding Forensic Group)	33	0.980	0.021	50.5%	50.0%	50.5%	1	48	47	1	97
	34	0.959	0.042	50.5%	50.0%	50.5%	2	47	46	2	97
	35	0.939	0.083	51.1%	57.1%	51.5%	3	46	44	4	97
	36	0.878	0.104	50.0%	45.5%	49.5%	6	43	43	5	97
	37	0.857	0.125	50.0%	46.2%	49.5%	7	42	42	6	97
	38	0.816	0.250	52.6%	57.1%	53.6%	9	40	36	12	97

(*table continues*)

Table 18 (*continued*):

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
Finger Oscillation Dominant Mean Raw (excluding Forensic Group - continued)	39	0.796	0.313	54.2%	60.0%	55.7%	10	39	33	15	97
	40	0.755	0.354	54.4%	58.6%	55.7%	12	37	31	17	97
	41	0.735	0.396	55.4%	59.4%	56.7%	13	36	29	19	97
	42	0.694	0.417	54.8%	57.1%	55.7%	15	34	28	20	97
	43	0.653	0.458	55.2%	56.4%	55.7%	17	32	26	22	97
	44	0.653	0.479	56.1%	57.5%	56.7%	17	32	25	23	97
BCT - Total Errors 51+ Errors = Impairment	35	1.000	0.039	40.7%	100.0%	42.1%	0	50	73	3	126
	37	0.940	0.066	39.8%	62.5%	41.3%	3	47	71	5	126
	38	0.920	0.079	39.7%	60.0%	41.3%	4	46	70	6	126
	41	0.880	0.092	38.9%	53.8%	40.5%	6	44	69	7	126
	42	0.880	0.105	39.3%	57.1%	41.3%	6	44	68	8	126
	43	0.860	0.132	39.4%	58.8%	42.1%	7	43	66	10	126
	44	0.840	0.158	39.6%	60.0%	42.9%	8	42	64	12	126
	45	0.840	0.171	40.0%	61.9%	43.7%	8	42	63	13	126
	46	0.840	0.184	40.4%	63.6%	44.4%	8	42	62	14	126
	47	0.780	0.211	39.4%	59.3%	43.7%	11	39	60	16	126
	50	0.780	0.237	40.2%	62.1%	45.2%	11	39	58	18	126
	51	0.760	0.237	39.6%	60.0%	44.4%	12	38	58	18	126
	52	0.760	0.263	40.4%	62.5%	46.0%	12	38	56	20	126
	53	0.740	0.263	39.8%	60.6%	45.2%	13	37	56	20	126
	54	0.740	0.276	40.2%	61.8%	46.0%	13	37	55	21	126
	55	0.740	0.289	40.7%	62.9%	46.8%	13	37	54	22	126
	56	0.740	0.303	41.1%	63.9%	47.6%	13	37	53	23	126
	58	0.740	0.342	42.5%	66.7%	50.0%	13	37	50	26	126
	59	0.720	0.355	42.4%	65.9%	50.0%	14	36	49	27	126
	60	0.680	0.355	41.0%	62.8%	48.4%	16	34	49	27	126
	61	0.660	0.355	40.2%	61.4%	47.6%	17	33	49	27	126
	62	0.640	0.368	40.0%	60.9%	47.6%	18	32	48	28	126
	64	0.620	0.382	39.7%	60.4%	47.6%	19	31	47	29	126

(*table continues*)

Table 18 (*continued*):

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
BCT - Total Errors	66	0.620	0.434	41.9%	63.5%	50.8%	19	31	43	33	126
(continued)	67	0.600	0.447	41.7%	63.0%	50.8%	20	30	42	34	126
	68	0.600	0.461	42.3%	63.6%	51.6%	20	30	41	35	126
	69	0.580	0.461	41.4%	62.5%	50.8%	21	29	41	35	126
	70	0.580	0.474	42.0%	63.2%	51.6%	21	29	40	36	126
	71	0.560	0.487	41.8%	62.7%	51.6%	22	28	39	37	126
	73	0.540	0.487	40.9%	61.7%	50.8%	23	27	39	37	126
WCST - Total Errors	5	0.980	0.039	40.2%	75.0%	41.3%	1	49	73	3	126
	7	0.940	0.053	39.5%	57.1%	40.5%	3	47	72	4	126
	8	0.920	0.079	39.7%	60.0%	41.3%	4	46	70	6	126
	9	0.760	0.092	35.5%	36.8%	35.7%	12	38	69	7	126
	10	0.760	0.132	36.5%	45.5%	38.1%	12	38	66	10	126
	11	0.700	0.158	35.4%	44.4%	37.3%	15	35	64	12	126
	12	0.660	0.158	34.0%	41.4%	35.7%	17	33	64	12	126
	13	0.620	0.211	34.1%	45.7%	37.3%	19	31	60	16	126
	14	0.500	0.224	29.8%	40.5%	33.3%	25	25	59	17	126
	15	0.500	0.250	30.5%	43.2%	34.9%	25	25	57	19	126
	16	0.460	0.263	29.1%	42.6%	34.1%	27	23	56	20	126
	17	0.440	0.276	28.6%	42.9%	34.1%	28	22	55	21	126
	18	0.440	0.316	29.7%	46.2%	36.5%	28	22	52	24	126
	19	0.420	0.316	28.8%	45.3%	35.7%	29	21	52	24	126
	20	0.380	0.342	27.5%	45.6%	35.7%	31	19	50	26	126
	21	0.360	0.368	27.3%	46.7%	36.5%	32	18	48	28	126
	22	0.360	0.395	28.1%	48.4%	38.1%	32	18	46	30	126
	23	0.340	0.408	27.4%	48.4%	38.1%	33	17	45	31	126
	24	0.320	0.421	26.7%	48.5%	38.1%	34	16	44	32	126
	26	0.300	0.421	25.4%	47.8%	37.3%	35	15	44	32	126
	27	0.280	0.434	24.6%	47.8%	37.3%	36	14	43	33	126
	28	0.260	0.447	23.6%	47.9%	37.3%	37	13	42	34	126

(*table continues*)

Table 18 (*continued*):

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
WCST - Total Errors (continued)	29	0.240	0.487	23.5%	49.3%	38.9%	38	12	39	37	126
	30	0.220	0.500	22.4%	49.4%	38.9%	39	11	38	38	126
	32	0.200	0.500	20.8%	48.7%	38.1%	40	10	38	38	126
	33	0.200	0.513	21.3%	49.4%	38.9%	40	10	37	39	126
	36	0.200	0.526	21.7%	50.0%	39.7%	40	10	36	40	126
	37	0.200	0.539	22.2%	50.6%	40.5%	40	10	35	41	126
	38	0.200	0.566	23.3%	51.8%	42.1%	40	10	33	43	126
	42	0.200	0.579	23.8%	52.4%	42.9%	40	10	32	44	126
	43	0.180	0.579	22.0%	51.8%	42.1%	41	9	32	44	126
	45	0.180	0.605	23.1%	52.9%	43.7%	41	9	30	46	126
	46	0.180	0.618	23.7%	53.4%	44.4%	41	9	29	47	126
	47	0.160	0.632	22.2%	53.3%	44.4%	42	8	28	48	126
	48	0.140	0.645	20.6%	53.3%	44.4%	43	7	27	49	126
	50	0.120	0.658	18.8%	53.2%	44.4%	44	6	26	50	126
	52	0.100	0.697	17.9%	54.1%	46.0%	45	5	23	53	126
	56	0.080	0.724	16.0%	54.5%	46.8%	46	4	21	55	126
	55	0.080	0.737	16.7%	54.9%	47.6%	46	4	20	56	126
	56	0.080	0.763	18.2%	55.8%	49.2%	46	4	18	58	126
	59	0.080	0.776	19.0%	56.2%	50.0%	46	4	17	59	126
	60	0.080	0.789	20.0%	56.6%	50.8%	46	4	16	60	126
	62	0.080	0.803	21.1%	57.0%	51.6%	46	4	15	61	126
	63	0.060	0.803	16.7%	56.5%	50.8%	47	3	15	61	126

Figure 17: *ROC Plot for WAIS-R NI Vocabulary T-Score*

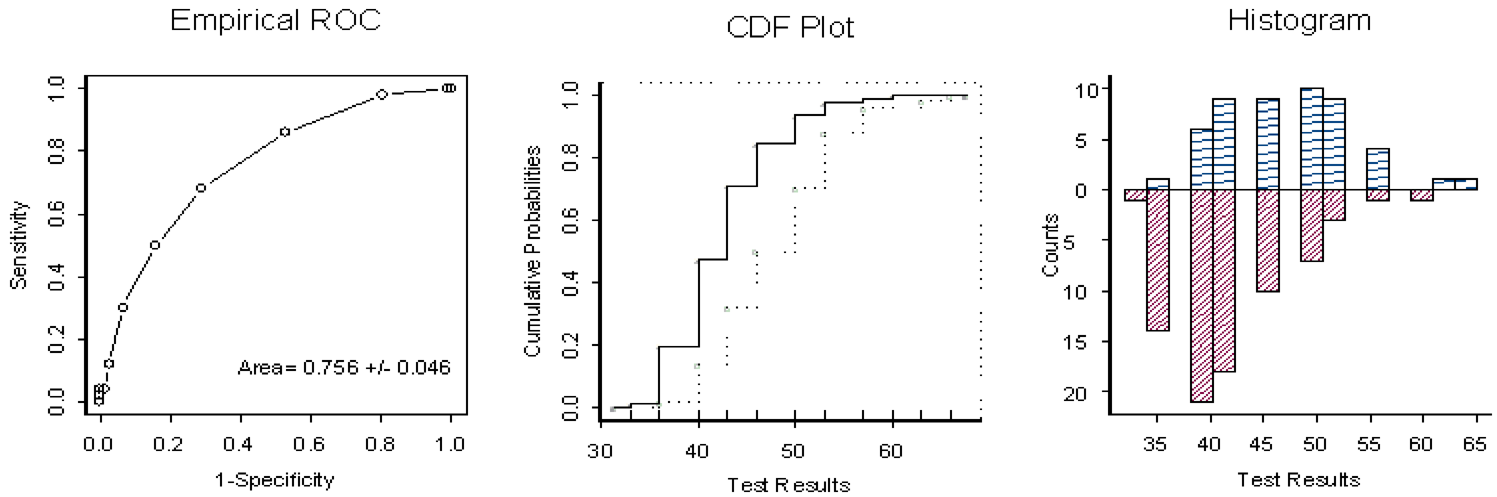


Figure 18: *ROC Plot for WAIS-R NI Information T-Score*

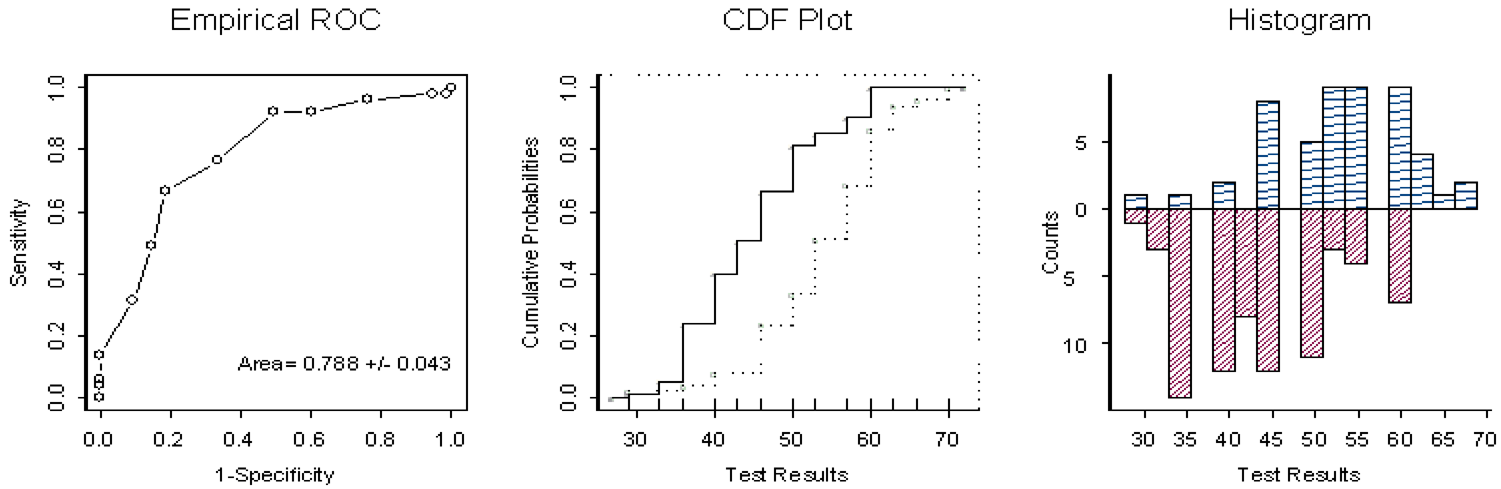


Figure 19: *ROC Plot for WAIS-R NI Similarities T-Score*

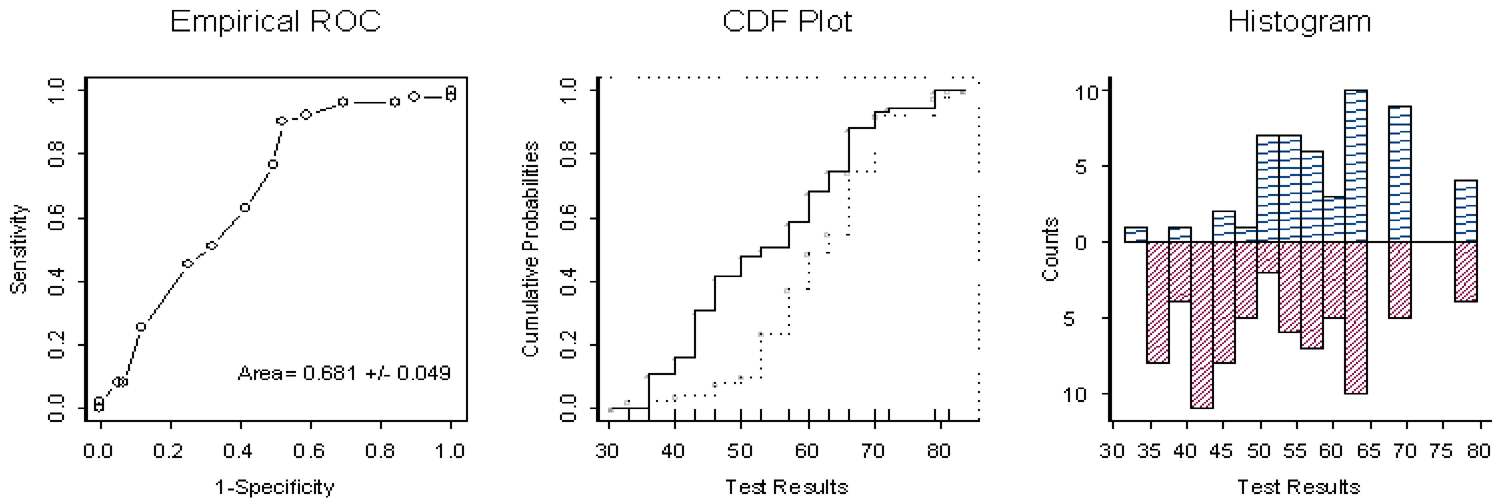


Figure 20: *ROC Plot for RMT by Group*

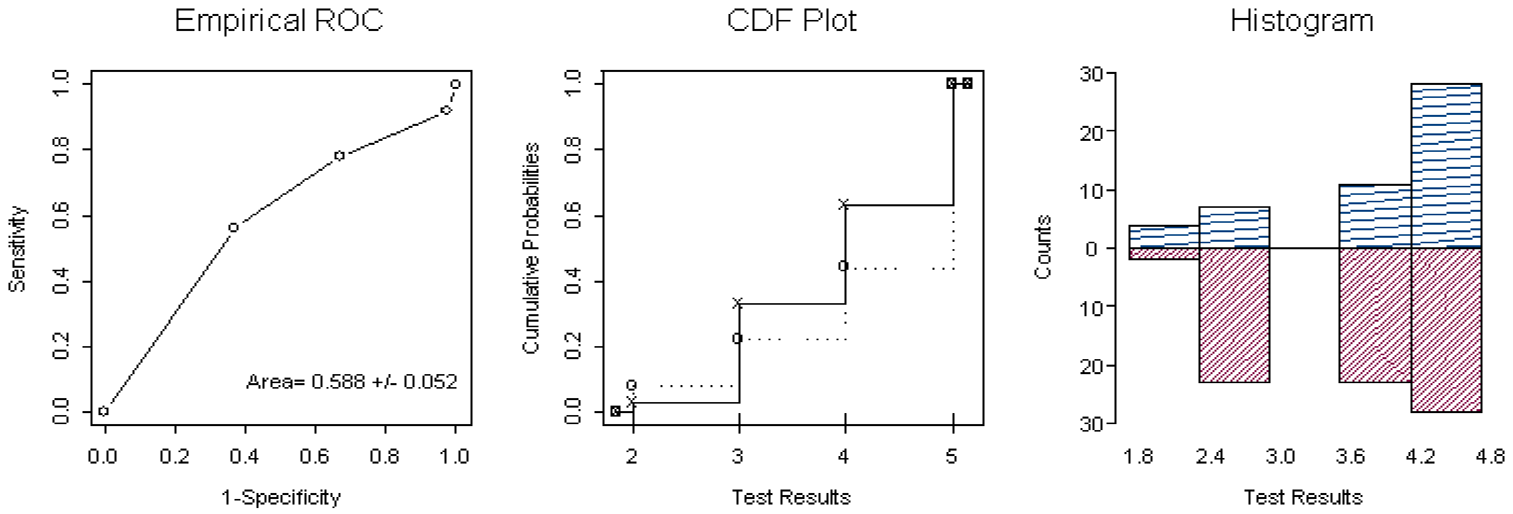


Figure 21: *ROC Plot for RMT by Total Items*

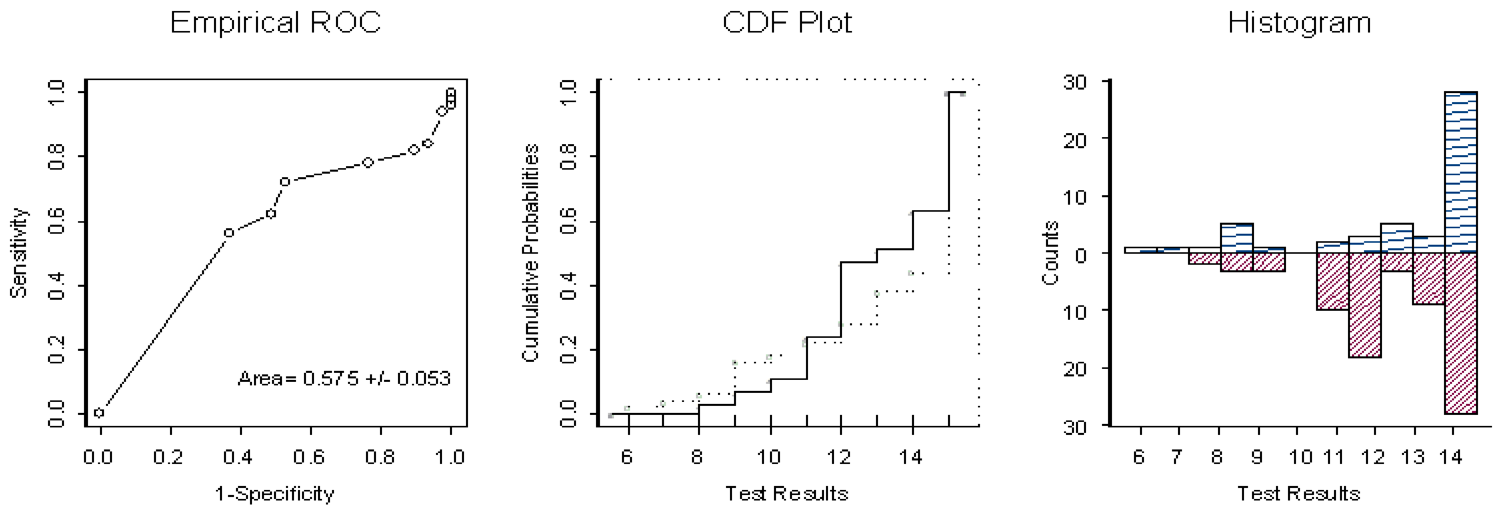


Figure 22: *ROC Plot for COWA Total Number Words*

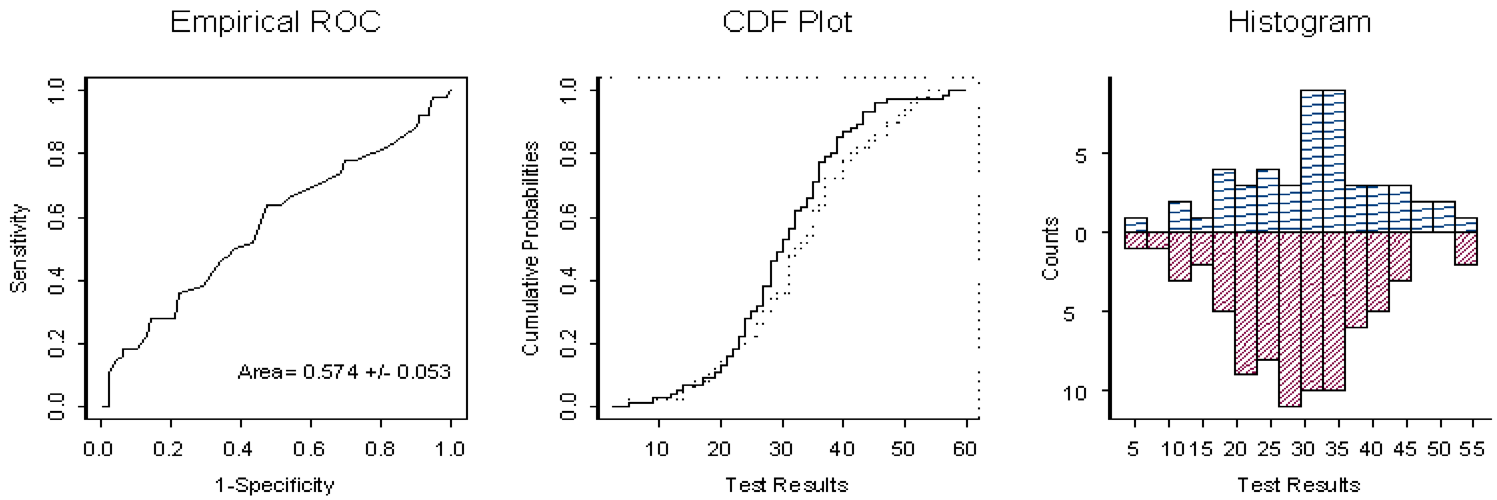


Figure 23: *ROC Plot for Dominant Hand Total Mean Finger Oscillation*

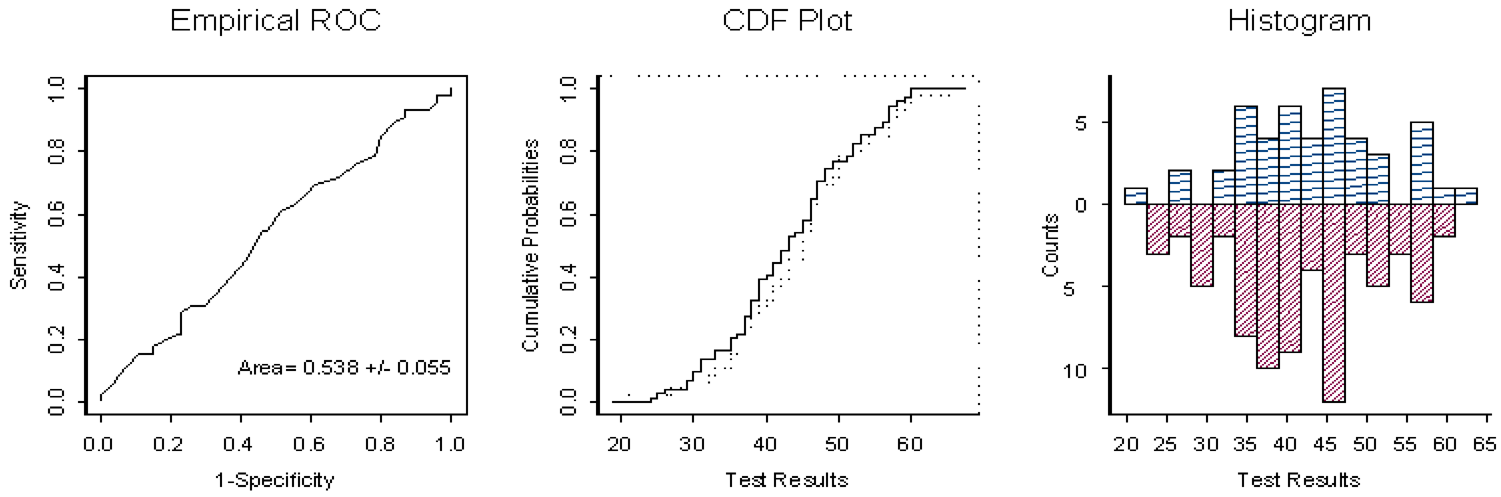


Figure 24: *ROC Plot for BCT Total Errors*

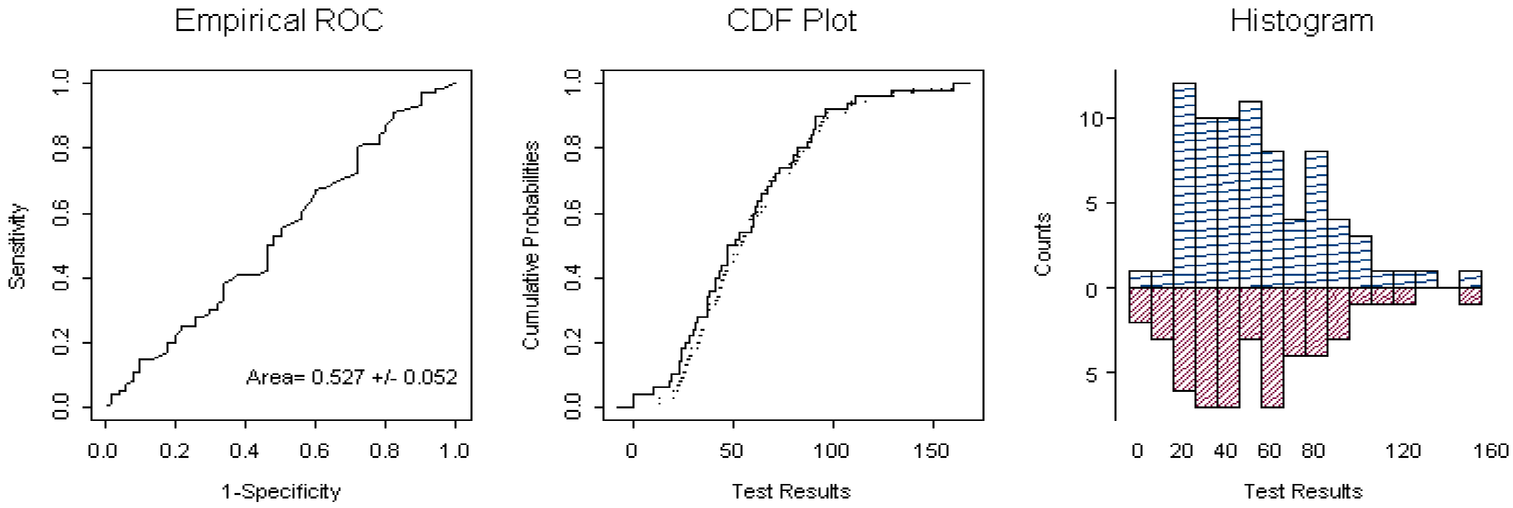


Figure 25: *ROC Plot for WCST Total Errors*

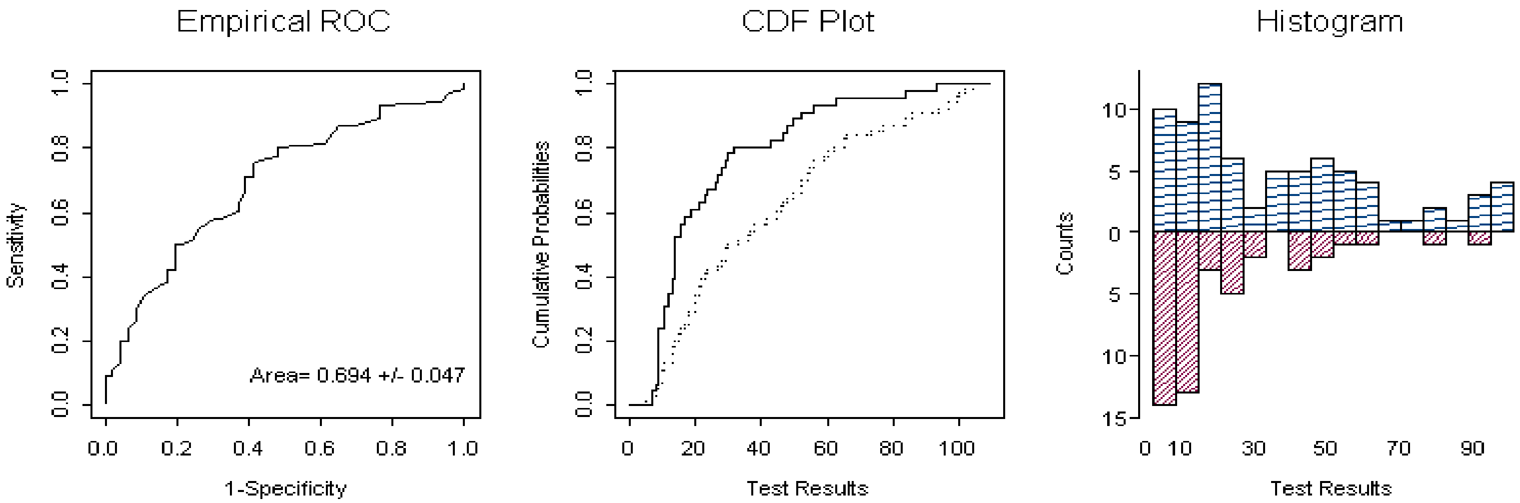


Table 19: *Efficacy of the WAIS-R NI Subtests - Simulating Group Compared to Both Head-Injured with and without Litigation Groups*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Vocabulary Subtest	33	68.0%	1.000	0.040	67.6%	100.0%
T-Score	36	74.7%	0.980	0.280	73.1%	87.5%
	40	74.7%	0.820	0.600	80.4%	62.5%
	43	64.0%	0.560	0.800	84.8%	47.6%
	46	53.3%	0.360	0.880	85.7%	40.7%
Information Subtest	33	70.7%	1.000	0.120	69.4%	100.0%
T-Score	36	73.3%	0.940	0.320	73.4%	72.7%
	40	77.3%	0.880	0.560	80.0%	70.0%
	43	78.7%	0.820	0.720	85.4%	66.7%
	46	73.3%	0.700	0.800	87.5%	57.1%
Similarities Subtest	33	68.0%	1.000	0.040	67.6%	100.0%
T-Score	36	76.0%	1.000	0.280	73.5%	100.0%
	40	80.0%	1.000	0.400	76.9%	100.0%
	43	85.3%	1.000	0.560	82.0%	100.0%
	46	88.0%	0.960	0.720	87.3%	90.0%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 20: *Efficacy of the RMT - Simulating Group Compared to Both Head-Injured with and without Litigation Groups*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Rey Total Items	6	65.3%	0.980	0.000	66.2%	0.0%
	7	64.0%	0.960	0.000	65.8%	0.0%
	8	62.7%	0.940	0.000	65.3%	0.0%
	9	62.7%	0.920	0.040	65.7%	20.0%
	10	66.7%	0.920	0.160	68.7%	50.0%
	11	74.7%	0.900	0.440	76.3%	68.8%
	12	78.7%	0.800	0.760	87.0%	65.5%
	13	74.7%	0.720	0.800	87.8%	58.8%
	14	68.0%	0.600	0.840	88.2%	51.2%
Rey Total Groups	2	62.7%	0.940	0.000	65.3%	0.0%
	3	82.7%	0.900	0.680	84.9%	77.3%
	4	68.0%	0.600	0.840	88.2%	51.2%
	5	33.3%	0.000	1.000	0.0%	33.3%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 21: *Efficacy of the COWA - Simulating Group Compared to Both Head-Injured with and without Litigation Groups*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
COWA Total Words	22	56.0%	0.760	0.160	64.4%	25.0%
	23	55.4%	0.720	0.208	65.5%	26.3%
	24	54.7%	0.680	0.280	65.4%	30.4%
	26	50.7%	0.620	0.280	63.3%	26.9%
	27	48.0%	0.560	0.320	62.2%	26.7%
	28	50.7%	0.540	0.440	65.9%	32.4%
	29	49.3%	0.520	0.440	65.0%	31.4%
	30	48.0%	0.500	0.440	64.1%	30.6%
	31	48.6%	0.468	0.520	64.7%	34.2%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 22: *Efficacy of Finger Oscillation - Simulating Group Compared to Both Head-Injured with and without Litigation Groups*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Total Mean Taps	28	69.3%	0.940	0.200	70.1%	62.5%
Dominant Hand	29	68.0%	0.900	0.240	70.3%	54.5%
	30	66.7%	0.880	0.240	69.8%	50.0%
	31	68.0%	0.860	0.320	71.7%	53.3%
	33	69.3%	0.860	0.360	72.9%	56.3%
	34	70.7%	0.860	0.400	74.1%	58.8%
	35	69.3%	0.820	0.440	74.5%	55.0%
	36	69.3%	0.800	0.480	75.5%	54.5%
	37	69.3%	0.780	0.520	76.5%	54.2%
	38	68.0%	0.700	0.640	79.5%	51.6%
	39	70.7%	0.680	0.760	85.0%	54.3%
	40	69.3%	0.640	0.800	86.5%	52.6%
	41	69.3%	0.620	0.840	88.6%	52.5%
	42	70.7%	0.620	0.880	91.2%	53.7%
	44	68.0%	0.560	0.920	93.3%	51.1%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 23: *Efficacy of the BCT - Simulating Group Compared to Both Head-Injured with and without Litigation Groups*

	Cutoff Score	Hit Rate	Sensitivity	Specificity	PPV	NPV
Total Number Errors	47	54.7%	0.560	0.520	70.0%	37.1%
	50	57.3%	0.560	0.600	73.7%	40.5%
	51	56.0%	0.540	0.600	73.0%	39.5%
	52	56.0%	0.520	0.640	74.3%	40.0%
	54	57.3%	0.520	0.680	76.5%	41.5%
	58	50.0%	0.740	0.342	42.5%	66.7%
	59	50.0%	0.720	0.355	42.4%	35.9%
	60	48.4%	0.680	0.355	41.0%	62.8%
	61	47.6%	0.660	0.355	40.0%	61.4%
	62	47.6%	0.640	0.368	40.0%	60.9%
	64	47.6%	0.640	0.382	39.7%	60.4%
	66	50.8%	0.620	0.434	41.9%	63.5%
	67	50.8%	0.600	0.447	41.7%	63.0%
	68	51.6%	0.600	0.461	42.3%	63.6%
	69	50.8%	0.580	0.461	41.4%	62.5%
	70	51.6%	0.580	0.474	42.0%	63.2%
	71	51.6%	0.560	0.487	41.8%	62.7%
	73	50.8%	0.540	0.487	40.9%	61.7%

NOTE: Hit Rate = Overall prediction accuracy (true positives + true negatives / total sample); Sensitivity = probability that a test makes a diagnosis when the disorder is present (true positives / (true positives + false negatives)); Specificity = probability that a test makes a negative diagnosis when the disorder is absent (true negatives / (true negatives + false positives)); PPV = Positive Predictive Value, likelihood that an individual has a disorder when a test predicts the presence of disease (true positives / (true positives + false positives)); NPV = Negative Predictive Value, likelihood that an individual does not have a disorder when a test predicts the absence of disease (true negatives / (true negatives + false negatives)).

Table 24: *Cutting Scores per Assessment Instrument - Simulating Group Compared to Head-Injured with and without Litigation Groups*

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
WAIS-R NI Vocabulary T-Score	33	1.000	0.040	67.6%	100.0%	68.0%	0	50	24	1	75
	36	0.980	0.280	73.1%	87.5%	74.7%	1	49	18	7	75
	40	0.820	0.600	80.4%	62.5%	74.7%	9	41	10	15	75
	43	0.560	0.800	84.8%	47.6%	64.0%	22	28	5	20	75
	46	0.360	0.880	85.7%	40.7%	53.3%	32	18	3	22	75
	50	0.180	0.920	81.8%	35.9%	42.7%	41	9	2	23	75
WAIS-R NI Information T-Score	33	1.000	0.120	69.4%	100.0%	70.7%	0	50	22	3	75
	36	0.940	0.320	73.4%	72.7%	73.3%	3	47	17	8	75
	40	0.880	0.560	80.0%	70.0%	77.3%	6	44	11	14	75
	43	0.820	0.720	85.4%	66.7%	78.7%	9	41	7	18	75
	46	0.700	0.800	87.5%	57.1%	73.3%	15	35	5	20	75
	50	0.520	0.880	89.7%	47.8%	64.0%	24	26	3	22	75
WAIS-R NI Similarities T-Score	33	1.000	0.040	67.6%	100.0%	68.0%	0	50	24	1	75
	36	1.000	0.280	73.5%	100.0%	76.0%	0	50	18	7	75
	40	1.000	0.400	76.9%	100.0%	80.0%	0	50	15	10	75
	43	1.000	0.560	82.0%	100.0%	85.3%	0	50	11	14	75
	46	0.960	0.720	87.3%	90.0%	88.0%	2	48	7	18	75
	50	0.900	0.800	90.0%	80.0%	86.7%	5	45	5	20	75
RMT - Total Items	6	0.980	0.000	66.2%	0.0%	65.3%	1	49	25	0	75
	7	0.960	0.000	65.8%	0.0%	64.0%	2	48	25	0	75
	8	0.940	0.000	65.3%	0.0%	62.7%	3	47	25	0	75
	9	0.920	0.040	65.7%	20.0%	62.7%	4	46	24	1	75
	10	0.920	0.160	68.7%	50.0%	66.7%	4	46	21	4	75
	11	0.900	0.440	76.3%	68.8%	74.7%	5	45	14	11	75
	12	0.800	0.760	87.0%	65.5%	78.7%	10	40	6	19	75
	13	0.720	0.800	87.8%	58.8%	74.7%	14	36	5	20	75
	14	0.600	0.840	88.2%	51.2%	68.0%	20	30	4	21	75

(*table continues*)

Table 24 (*continued*)

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
RMT - Total Groups	2	0.940	0.000	65.3%	0.0%	62.7%	3	47	25	0	75
	3	0.900	0.680	84.9%	77.3%	82.7%	5	45	8	17	75
	4	0.600	0.840	88.2%	51.2%	68.0%	20	30	4	21	75
	5	0.000	1.000	- - - -	33.3%	33.3%	50	0	0	25	75
COWA - Total Raw Words	20	0.820	0.160	66.1%	30.8%	60.0%	9	41	21	4	75
	21	0.780	0.160	65.0%	26.7%	57.3%	11	39	21	4	75
	22	0.760	0.160	64.4%	25.0%	56.0%	12	38	21	4	75
	23	0.720	0.208	65.5%	26.3%	55.4%	14	36	19	5	74
	24	0.680	0.280	65.4%	30.4%	54.7%	16	34	18	7	75
	26	0.620	0.280	63.3%	26.9%	50.7%	19	31	18	7	75
	27	0.560	0.320	62.2%	26.7%	48.0%	22	28	17	8	75
	28	0.540	0.440	65.9%	32.4%	50.7%	23	27	14	11	75
	29	0.520	0.440	65.0%	31.4%	49.3%	24	26	14	11	75
	30	0.500	0.440	64.1%	30.6%	48.0%	25	25	14	11	75
	31	0.468	0.520	64.7%	34.2%	48.6%	25	22	12	13	72
	32	0.420	0.600	67.7%	34.1%	48.0%	29	21	10	15	75
	33	0.400	0.600	66.7%	33.3%	46.7%	30	20	10	15	75
	34	0.380	0.640	67.9%	34.0%	46.7%	31	19	9	16	75
	35	0.380	0.680	70.4%	35.4%	48.0%	31	19	8	17	75
Finger Oscillation Dominant Mean Raw	27	0.940	0.160	69.1%	57.1%	68.0%	3	47	21	4	75
	28	0.940	0.200	70.1%	62.5%	69.3%	3	47	20	5	75
	29	0.900	0.240	70.3%	54.5%	68.0%	5	45	19	6	75
	30	0.880	0.240	69.8%	50.0%	66.7%	6	44	19	6	75
	31	0.860	0.320	71.7%	53.3%	68.0%	7	43	17	8	75
	33	0.860	0.360	72.9%	56.3%	69.3%	7	43	16	9	75
	34	0.860	0.400	74.1%	58.8%	70.7%	7	43	15	10	75
	35	0.820	0.440	74.5%	55.0%	69.3%	9	41	14	11	75
	36	0.800	0.480	75.5%	54.5%	69.3%	10	40	13	12	75
	37	0.780	0.520	76.5%	54.2%	69.3%	11	39	12	13	75

(*table continues*)

Table 24 (*continued*)

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
Finger Oscillation (continued)	38	0.700	0.640	79.5%	51.6%	68.0%	15	35	9	16	75
	39	0.680	0.760	85.0%	54.3%	70.7%	16	34	6	19	75
	40	0.640	0.800	86.5%	52.6%	69.3%	18	32	5	20	75
	41	0.620	0.840	88.6%	52.5%	69.3%	19	31	4	21	75
	42	0.620	0.880	91.2%	53.7%	70.7%	19	31	3	22	75
	44	0.560	0.920	93.3%	51.1%	68.0%	22	28	2	23	75
	45	0.460	0.960	95.8%	47.1%	62.7%	27	23	1	24	75
	46	0.400	1.000	100.0%	45.5%	60.0%	30	20	0	25	75
	47	0.340	1.000	100.0%	43.1%	56.0%	33	17	0	25	75
BCT - Total Errors >51 Errors = Impairment	36	0.720	0.280	66.7%	33.3%	57.3%	14	36	18	7	75
	37	0.680	0.320	66.7%	33.3%	56.0%	16	34	17	8	75
	38	0.660	0.360	67.3%	34.6%	56.0%	17	33	16	9	75
	41	0.620	0.360	66.0%	32.1%	53.3%	19	31	16	9	75
	44	0.620	0.440	68.9%	36.7%	56.0%	19	31	14	11	75
	45	0.620	0.480	70.5%	38.7%	57.3%	19	31	13	12	75
	46	0.600	0.480	69.8%	37.5%	56.0%	20	30	13	12	75
	47	0.560	0.520	70.0%	37.1%	54.7%	22	28	12	13	75
	50	0.560	0.600	73.7%	40.5%	57.3%	22	28	10	15	75
	51	0.540	0.600	73.0%	39.5%	56.0%	23	27	10	15	75
	52	0.520	0.640	74.3%	40.0%	56.0%	24	26	9	16	75
	54	0.520	0.680	76.5%	41.5%	57.3%	24	26	8	17	75
	58	0.500	0.760	80.6%	43.2%	58.7%	25	25	6	19	75
	62	0.460	0.760	79.3%	41.3%	56.0%	27	23	6	19	75
	64	0.440	0.800	81.5%	41.7%	56.0%	28	22	5	20	75
	66	0.400	0.840	83.3%	41.2%	54.7%	30	20	4	21	75
	67	0.380	0.840	82.6%	40.4%	53.3%	31	19	4	21	75
	71	0.360	0.840	81.8%	39.6%	52.0%	32	18	4	21	75
	73	0.340	0.840	81.0%	38.9%	50.7%	33	17	4	21	75
	82	0.320	0.840	80.0%	38.2%	49.3%	34	16	4	21	75
	83	0.300	0.840	78.9%	37.5%	48.0%	35	15	4	21	75

60 0.48

(*table continues*)

Table 24 (*continued*)

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
BCT - Total Errors (continued)	87	0.280	0.840	77.8%	36.8%	46.7%	36	14	4	21	75
	88	0.280	0.920	87.5%	39.0%	49.3%	36	14	2	23	75
	89	0.260	0.960	92.9%	39.3%	49.3%	37	13	1	24	75
	90	0.240	1.000	100.0%	39.7%	49.3%	38	12	0	25	75
	91	0.220	1.000	100.0%	39.1%	48.0%	39	11	0	25	75
	95	0.200	1.000	100.0%	38.5%	46.7%	40	10	0	25	75
	96	0.184	1.000	100.0%	38.5%	45.9%	40	9	0	25	74
	106	0.160	1.000	100.0%	37.3%	44.0%	42	8	0	25	75
	107	0.140	1.000	100.0%	36.8%	42.7%	43	7	0	25	75
	111	0.120	1.000	100.0%	36.2%	41.3%	44	6	0	25	75
	116	0.100	1.000	100.0%	35.7%	40.0%	45	5	0	25	75
	129	0.080	1.000	100.0%	35.2%	38.7%	46	4	0	25	75
	130	0.060	1.000	100.0%	34.7%	37.3%	47	3	0	25	75
	140	0.040	1.000	100.0%	34.2%	36.0%	48	2	0	25	75
	160	0.000	1.000	- - - -	33.3%	33.3%	50	0	0	25	75
WCST - Total Errors	7	0.880	0.000	63.8%	0.0%	58.7%	6	44	25	0	75
	8	0.860	0.040	64.2%	12.5%	58.7%	7	43	24	1	75
	9	0.780	0.040	61.9%	8.3%	53.3%	11	39	24	1	75
	10	0.720	0.040	60.0%	6.7%	49.3%	14	36	24	1	75
	11	0.660	0.040	57.9%	5.6%	45.3%	17	33	24	1	75
	12	0.620	0.040	56.4%	5.0%	42.7%	19	31	24	1	75
	13	0.560	0.120	56.0%	12.0%	41.3%	22	28	22	3	75
	14	0.520	0.120	54.2%	11.1%	38.7%	24	26	22	3	75
	16	0.480	0.120	52.2%	10.3%	36.0%	26	24	22	3	75
	17	0.460	0.120	51.1%	10.0%	34.7%	27	23	22	3	75
	18	0.420	0.160	50.0%	12.1%	33.3%	29	21	21	4	75
	20	0.360	0.160	46.2%	11.1%	29.3%	32	18	21	4	75
	21	0.320	0.160	43.2%	10.5%	26.7%	34	16	21	4	75
	22	0.320	0.240	45.7%	15.0%	29.3%	34	16	19	6	75
	23	0.300	0.240	44.1%	14.6%	28.0%	35	15	19	6	75

(*table continues*)

Table 24 (*continued*)

Assessment Instrument	Cutting Score	Sensitivity	Specificity	PPV	NPV	Hit Rate	FN	TP	FP	TN	N
WCST - Total Errors (continued)	24	0.280	0.240	42.4%	14.3%	26.7%	36	14	19	6	75
	26	0.260	0.240	40.6%	14.0%	25.3%	37	13	19	6	75
	27	0.240	0.240	38.7%	13.6%	24.0%	38	12	19	6	75
	28	0.180	0.280	33.3%	14.6%	21.3%	41	9	18	7	75
	29	0.180	0.320	34.6%	16.3%	22.7%	41	9	17	8	75
	30	0.160	0.320	32.0%	16.0%	21.3%	42	8	17	8	75
	33	0.160	0.360	33.3%	17.6%	22.7%	42	8	16	9	75
	36	0.160	0.360	33.3%	17.6%	22.7%	42	8	16	9	75
	38	0.160	0.400	34.8%	19.2%	24.0%	42	8	15	10	75
	43	0.140	0.400	31.8%	18.9%	22.7%	43	7	15	10	75
	45	0.140	0.440	33.3%	20.4%	24.0%	43	7	14	11	75
	47	0.120	0.440	30.0%	20.0%	22.7%	44	6	14	11	75
	48	0.100	0.440	26.3%	19.6%	21.3%	45	5	14	11	75
	52	0.080	0.500	23.5%	22.0%	22.4%	46	4	13	13	76
	54	0.060	0.500	18.8%	21.7%	21.1%	47	3	13	13	76
	56	0.040	0.500	13.3%	21.3%	19.7%	48	2	13	13	76
	62	0.040	0.560	15.4%	22.6%	21.3%	48	2	11	14	75
	63	0.020	0.560	8.3%	22.2%	20.0%	49	1	11	14	75
	65	0.020	0.600	9.1%	23.4%	21.3%	49	1	10	15	75
	66	0.020	0.640	10.0%	24.6%	22.7%	49	1	9	16	75
	73	0.020	0.680	11.1%	25.8%	24.0%	49	1	8	17	75

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.

Axelrod, B.N., Goldman, R.S., Heaton, R.K., Curtiss, G., Thompson, L.L., Chelune, G.J., & Kay, G.G. (1996). Discriminability of the Wisconsin Card Sorting Test using the standardization sample. *Journal of Clinical and Experimental Neuropsychology, 18*(3), 338-342.

Barth, J.T., Gideon, D.A., Sciara, S.,D., Hulsey, P.H., & Anchor, K.N. (1986). Forensic aspects of mild head injury. *Journal of Head Trauma Rehabilitation, 1*, 63-70.

Bash, I. Y., & Alpert, M. (1980). The determination of malingering. *Annals of the New York Academy of Science, 347,* 86-99.

Bechtoldt, H.P., Benton, A.L., & Fogel, M.L. (1962). An application of factor analysis in neuropsychology. *Psychological Record, 12*, 147-156.

Benton, A. L., & Spreen, O. (1961). Visual Memory Test. *Archives of General Psychiatry*, 4, 79-83.

Berg, E.A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology, 29,* 15-22.

Bernard, L. C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology, 12*(5), 715-728.

Bernard, L. C., & Fowler, W. (1990). Assessing the validity of memory complaints: Performance of brain-damaged and normal individuals on Rey's task to detect malingering. *Journal of Clinical Psychology, 46,* 432-436.

Bernard, L.C., Houston, W., & Natoli, L. (1993). Malingering on neuropsychological memory tests: Potential objective indicators. *Journal of Clinical Psychology, 49*(1), 45-53.

Bernard, L.C., McGrath, M.J., and Houston, W. (1996). The differential effects of simulating malingering, closed head injury, and other CNS pathology on the Wisconsin Card Sorting Test: Support for the "Pattern of Performance" hypothesis. *Archives of Clinical Neuropsychology, 11,* 231-245.

Berry, D.T.R., Lamb, D.G., Wetter, M.W., Baer, R.A., & Widiger, T.A. (1994). Ethical considerations in research on coached malingering. *Psychological Assessment, 6*(1), 16-17.

Bickert, W. T., Meyer, R.G., & Connell, D.K. (1991). The symptom validity technique as a measure of feigned short-term memory deficit. *American Journal of Forensic Psychology, 9*(2), 3-11.

Bigler, E.D. (1986). Forensic Issues in Neuropsychology. In D.W. Wedding, A M Horton, Jr., & J. Webster (Eds.), *The Neuropsychology Handbook: Behavioral and clinical perspectives.* (pp. 526-547). New York: Springer Publishing Company.

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

Binder, L. M. (1992a). Deception and malingering. In A.E. Puente & R.J. McCaffrey (Eds.), *Handbook of neuropsychological assessment: A biopsychosocial perspective*. (pp. 353-374). New York: Plenum Press.

Binder, L. M. (1992b). Malingering detected by forced choice testing of memory and tactile sensation: A case report. *Archives of Clinical Neuropsychology, 7*(2), 155-163.

Binder, L. M., & Pankratz, L. (1987). Neuropsychological evidence of a factitious memory complaint. *Journal of Clinical and Experimental Neuropsychology, 9,* 167-171.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*, 175-181.

Binks, P.G., Gouvier, W.D., & Waters, W.F. (1997). Malingering detection with the Dot Counting Test. *Archives of Clinical Neuropsychology, 12*(1), 41-46.

Bolter, J. F. (1992). *Item error frequencies for the Halstead category test: A cross validation study on a performance validity check*. Unpublished manuscript, Dept. of Psychiatry, Silas B. Hays Army Community Hospital, Fort Ord, CA.

Boone, K.B., Savodnik, I., Ghaffarian, S., Lee, A., Freeman, D., & Berman, N.G. (1995). Rey 15-Item Memorization and Dot Counting scores in a stress" claim worker's compensation population: Relationship to personality (MCMI) scores. *Journal of Clinical Psychology, 5*(3), 457-463.

Brandt, J. (1988). Malingered amnesia. In R. Rogers (ed.), *Clinical Assessment of Malingering and Deception*, (pp.65-83). New York: Guilford.

Bruhn, A.R. & Reed, M.R. (1975). Simulation of brain damage on the Bender-Gestalt Test by college subjects. *Journal of Personality Assessment, 39*(3), 244-255.

Butler, M., Retzlaff, P., & Vanderploeg, R. (1991). Neuropsychological test usage. *Professional Psychology: Research and Practice, 22*. 510-521.

Cohen, J. (2001). A Power Primer. *Psychological Bulletin, 112*(1), 155-159.

Cradock, M.M., & Gfeller, J.D. (1996). The ability of simulating subjects to accurately feign post-concussion symptomatology. *Archives of Clinical Neuropsychology, Abstracts from the 15th Annual Meeting of the National Academy of Neuropsychology*, *11*(5), 379.

Davis, H.P., King, J.H., Klebe, K.J., Bajszar, G. Jr., Bloodworth, M.R., & Wallick, S.L. (1997). The detection of simulated malingering using a computerized priming test. *Archives of*

Clinical Neuropsychology, 12(2), 145-153.

Dodrill, C.B. (1979). Sex differences on the Halstead-Reitan Neuropsychological Battery and on other neuropsychological measures. *Journal of Clinical Psychology,* 35. 236-241.

Faul, F. & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS (Computer program). Bonn, FRG: Bon University, Department of Psychology.

Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56*(4), 578-582.

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

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

Frederick, R.I., & Foster, Jr., H.G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. *Journal of Consulting and Clinical Psychology, 3*(4), 596-602.

Gallucci, N.T. (1984). Prediction of dissimulation on the MMPI in a clinical field setting. *Journal of Consulting and Clinical Psychology, 52*, 917-918.

Goebel, R. A. (1983). Detection of faking on the Halstead-Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology, 39*, 731-742.

Goldberg, J.O., & Miller, H.R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints.

Journal of Clinical Psychology, 42(5), 792-795.

Golden, C.J. (1981). A standardized version of Luria's neuropsychological tests. S. Filskov and TJ. Boll (Eds.). *Handbook of Clinical Neuropsychology*. New York: Wiley-Interscience.

Gorman, W. F. (1984). Neurological malingering. *Behavioral Sciences and the Law, 2*(1), 67-73.

Green, P., Iverson, G. & Allen, L. (1999). Detecting Malingering in Head Injury Litigation with the Word Memory Test. *Brain Injury, 13*(10), 813-819.

Green, P., Rohling, M.L., Lees-Haley, P.R. & Allen L.M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*(12), 1045-1060.

Greiffenstein, M.F., Baker, W.J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*(3), 218-224.

Greiffenstein, M.F., Baker, W.J., & Gola, T. (1996). Comparison of multiple scoring methods for Rey's malingered amnesia measures. *Archives of Clinical Neuropsychology, 11*(4), 283-293.

Gudjonsson, G. H., & Shackleton, H. (1986). The pattern of scores on Raven's Matrices during "faking bad" and "non faking" performances. *British Journal of Clinical Psychology*, 25, 35-41.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test Manual*. Florida: Psychological Assessment Resources, Inc.

Heaton, R. K., Smith, Jr. H. H., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical*

Psychology, 46(5), 892-900.

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11*(6), 967-974.

Holtz, J.L., Gearhart, L.P., & Watson, C.G. (1996). Comparability of scores on projector- and booklet-administered forms of the Category Test in brain-impaired veterans and controls. *Neuropsychology, 10*(2), 194-196.

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 Behavior, 15*(6), 667-676.

Iverson, G.L. & Binder, L.M. (2002). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of head Trauma Rehabilitation,* 15, 829-858.

Iverson, G.L., & Tulsky, D.S. (2003). Detecting malingering on the WAIS-III: Unusual Digit Span performance patterns in the normal population and in clinical groups. *Archives of Clinical Neuropsychology, 18*(1), 1-9.

Johnson E.E., Hamer, R., Nora, R.M., & Tan, B. (1997). The Lie/Bet questionnaire for screening pathological gamblers. *Psychological Reports, 80*(1), 83-88.

Johnson, J. L., & Lesniak-Karpiak, K. (1997). The effect of warning on malingering on memory and motor tasks in college samples. *Archives of Clinical Neuropsychology, 12*(3), 231-238.

Johnstone, B. Holland, D., & Hewett, J.E. (1997). The construct validity of the Category Test: Is it a measure of reasoning or intelligence? *Psychological Assessment, 9*(1), 288-22.

Kaplan, E., Fein, D., Morris, R., & Delis, D.C. (1991). *WAIS-R as a Neuropsychological Instrument.* Toronto: The Psychological Corporation, Harcourt Brace Jovanovich, Inc.

Kirk, R.E. (1990). In R.E. Kirk (Ed.). *Statistics: An introduction*. (3rd ed.). Fort Worth: Holt, Rinehart and Winston, Inc.

Knight, J.A., Webster, J., Goetsch, V., Malloy, P.F., and Greve, K. (1986). *Utilizing the Wisconsin Card Sorting Test to detect malingered performance.* Unpublished manuscript presented at the 94th annual Meeting of the American Psychological Association, Washington, DC.

Kolb, B., & Whishaw, I.Q. (1996). In R.C. Atkinson, G. Lindzey, & R.F. Thompson (Eds.). *Fundamentals of Human Neuropsychology*. 4th ed. New York: W.H. Freeman and Company.

Labarge, A.S., McCaffrey, R.M. & Brown, T.A. (2003). Neuropsychologists' abilities to determine the predictive value of diagnostic tests. *Archives of Clinical Neuropsychology, 18*(2), 165-175.

Larrabee, G.J. (1990). Cautions in the use of neuropsychological evaluation in legal settings. *Neuropsychology, 4,* 239-247.

Larrabee, G.J. (2002). Neuropsychological performance patterns indicative of malingering. Presented at the 30th Annual Meeting of the International Neuropsychological Society, Toronto, Canada.

Lee, G.P., Loring, D.W., & Martin, R.C. (1992). Rey's 15-Item Visual memory Test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment, 4*(1), 43-46.

Lees-Haley, P.R. (1990). Provisional normative data for a credibility scale for assessing personal injury claimants. *Psychological Reports, 66,* 1355-1360.

Lezak, M. D. (1983). *Neuropsychological Assessment,* (pp. 615-622). New York: Oxford

Press.

Macmillan, N.A. & Creelman, C.D. (1991). *Detection Theory: A User's Guide*. New York: Cambridge University Press.

Mensch, A. J., & Woods, D. J. (1986). Patterns of feigning brain damage on the LNNB. *International Journal of Clinical Neuropsychology, 8*, 59-63.

Metz C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4), 283-298.

McKinzey, R. K., Podd, M. H., Krehbiel, M. A., & Raven, J. (1999). Detection of malingering on the Raven's Standard Progressive Matrices: A cross-validation. *British Journal of Clinical Psychology*, 38, 435-439.

McKinzey, R.K., & Russell, E.W. (1997). Detection of malingering on the Halstead-Reitan Battery: A cross-validation. *Archives of Clinical Neuropsychology, 12*(6), 585-589.

Mittenberg, W. (1990). *Neuropsychological assessment of malingering*. Unpublished manuscript, Nova University School of Psychology, Fort Lauderdale.

Mittenberg, W., D'Attilio, J., Gage, R., & Bass, A. (1990). *Malingered symptoms following head trauma: Memory and post-concussion syndromes*. Paper presented at the 18th meeting of the International Neuropsychological Society, Orlando, FL.

Mittenberg, W., Rotholc, A., Russell, E., & Heilbronner, R. (1996). Identification of malingered head injury on the Halstead-Reitan Battery. *Archives of Clinical Neuropsychology, 11*(4), 271-281.

Moses, J. A., Golden, C. J., Wilkening, G. N., McKay, S. E., & Ariel, R. (1983). *Interpretation of the Luria-Nebraska Neuropsychological Battery* (Vol. 2). New York: Grune &

Stratton.

Ozonoff, S. (1995). Reliability and Validity of the Wisconsin Card Sorting Test in Studies of Autism. *Neuropsychology, 9*(4), 491-500.

Pankratz, L. (1979). Symptom validity testing and symptom retaining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology, 47*(2), 409-410.

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills, 57*, 367-372.

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

Prigatano, G.P., & Amin, K. (1993). Digit Memory Test: Unequivocal Cerebral Dysfunction and Suspected Malingering. *Journal of Clinical and Experimental Neuropsychology*, 15(4), 537-546.

Prigatano, G.P., Samson, I. Lamb, D.G., & Bortz, J.J. (1997). Suspected Malingering and the Digit Memory Test: A replication and extension. *Archives of Clinical Neuropsychology, 12*(7), 609-619.

Powell, K.E. (1991). *The malingering of schizophrenia*. Unpublished doctoral dissertation, University of South Carolina, Columbia.

Psychological Corporation. (1991). *WAIS-R as a neuropsychological instrument*. [Brochure]. Kaplan, E., Fein, D., Morris, R., & Delis, DC: Authors.

Reitan, R.M., & Wolfson, D. (1985a). Theory and rationale of the Halstead-Reitan Neuropsychological Test Battery. In R.M. Reitan & D. Wolfson's (eds.), *The Halstead-Reitan*

Neuropsychological Test Battery: Theory and clinical interpretation. (pp.1-13). Arizona: Neuropsychology Press.

Reitan, R.M., & Wolfson, D. (1985b). *Neuroanatomy and Neuropathology: A clinical guide for neuropsychologists*. Arizona: Neuropsychology Press.

Reitan, R.M., & Wolfson, D. (1986). *Traumatic Brain Injury: Volume I; Pathophysiology and neuropsychological evaluation*. Arizona: Neuropsychology Press.

Reitan, R.M., & Wolfson, D. (1993). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*. (2nd ed.). Arizona: Neuropsychology Press.

Resnick, P.J. (1988). Malingering of post-traumatic disorders. In R. Rogers (ed.), *Clinical Assessment of Malingering and Deception* (pp. 84-103). New York: Guilford.

Resnick, P.J. (1992). *The detection of malingered mental illness*. Unpublished manuscript. Presented at the American Psychology-Law Society, Division 41 Mid-Year Conference, San Diego, California.

Rey, A. (1964). L'examen clinique en psychologie. Paris: Presses Universitaires de France.

Robinson, A.L., Heaton, R.K., Lehman, R.W,. & Stilson, D.W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting and Clinical Psychology, 48*. 605-614.

Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law, 2*, 93-111.

Rogers, R. (1986). *Conducting Insanity Evaluations*. New York: Van Nostrand Reinhold.

Rogers, R. (1988a). Introduction. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception*, (pp. 1-9). New York: Guilford.

Rogers, R. (1988b). Researching dissimulation. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception*, (pp. 309-327). New York: Guilford.

Rogers, R. (1990). Models of feigned mental illness. *Professional Psychology: Research and Practice, 21*(3), 182-188.

Rogers, R. (1995). The nature of diagnostic and structured interviewing. In R. Rogers (Ed.), *Diagnostic and Structured Interviewing: A handbook for psychologists*, (pp. 1-27). Odessa: Psychological Assessment Resources, Inc.

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.

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*(4), 329-338.

Sackett, D.L., Richardson, W.S., Rosenberg, W., & Haynes, R.B. (1997). *Evidence-based medicine: how to practice and teach EBM*. New York: Churchill Livingston.

Schretlen, D., Brandt, J., & Krafft, L, & Van Gorp, W. (1991). Some caveats in using the Rey 15-item Memory Test to detect malingered amnesia. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*(4), 667-672.

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

Slick, D., Hopp, G., Strauss, E., Fox, D., Pinch, D., & Stickgold, K. (1996). Effects of prior testing with the WAIS-R NI on subsequent retest with the WAIS-R. *Archives of Clinical*

Neuropsychology, 11(2), 123-130.

Simon, M.J. (1994). The use of the Rey Memory Test to assess malingering in criminal defendants. *Journal of Clinical Psychology, 50*(6), 913-917.

Suhr, J.A., & Boyer, D. (1999). Use of the Wisconsin Card Sorting Test in the Detection of Malingering in Student Simulator and Patient Samples. Journal of Clinical and Experimental Neuropsychology, 21, 701-708.

Sweet, J.J. (1999). Malingering: Differential Diagnosis. In J. J. Sweet (Ed.), *Forensic Neuropsychology: Fundamentals and practice*. Swets & Zeitlinger.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-93.

Tombaugh, T.N. (1996). The Test of Memory Malingering. Toronto, Canada: Multi-Health Systems

Tabachnick, B.G. & Fidell, L.S. (2001). *Using Multivariate Statistics*. Allyn and Bacon: Boston.

Trueblood, W., & Binder, L.M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingers. *Archives of Clinical Neuropsychology, 12*(1), 13-27.

Vogt, P.W. (1993). In P.W. Vogt (Ed.), *Dictionary of Statistics and Methodology: A nontechnical guide for the social sciences*. Newbury Park, CA: Sage Publications, Inc.

Wasyliw, O. E., & Golden, C. J. (1985). Neuropsychological evaluation in the assessment of personal injury. *Behavioral Sciences and the Law, 3,* 149-164.

Youngjohn, J. R. (1991). Malingering of neuropsychological impairment: An assessment strategy. *A Journal for the Expert Witness, the Trial Attorney, the Trial Judge, 4*, 29-32.

Ziskin, J. (1984). Malingering of psychological disorders. *Behavioral Sciences and the Law, 2,* 39-49.