DETECTION OF MALINGERING ON RAVEN'S STANDARD PROGRESSIVE MATRICES AND THE BOOKLET CATEGORY TEST

DISSERTATION

Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

William Charles Isler, III, B.A., M.A.

Denton, Texas

December , 1997

Isler, William Charles III, Detection of malingering on Raven's Standard Progressive Matrices and the Booklet Category Test. Doctor of Philosophy (Clinical Psychology), December, 1997, 139 pp., 30 tables, references, 117 titles.

The capacity of Raven's Standard Progressive Matrices (SPM) and the Booklet Category Test (BCT) to discriminate between groups of brain-injured, simulated malingering, and normal participants was investigated in this study. Exploratory analyses were also conducted to examine the differences between groups categorized as sophisticated and naive fakers. Clinical decision rules and discriminant function analyses were utilized to identify malingerers. Clinical decision rules ranged in hit rates from 41% to 78%, in sensitivity from 2% to 100%, and in specificity from 86% to 100%. Discriminant functions ranged in hit rates from 81% to 86%, in sensitivity from 68% to 73% and in specificity from 82% to 87%. Overall, the least helpful detection method examined was below chance responding on either measure, while the most efficient was gross errors for SPM.

DETECTION OF MALINGERING ON RAVEN'S STANDARD PROGRESSIVE MATRICES AND THE BOOKLET CATEGORY TEST

DISSERTATION

Presented to the Graduate Council of the

University of North Texas in Partial

Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

William Charles Isler, III, B.A., M.A.

Denton, Texas

December , 1997

ACKNOWLEDGMENTS

The author would like to thank the numerous people and agencies which made this research possible. The University of North Texas awarded a small grant which provided partial funding for incentive pay for simulation participants. I would also like to thank the committee members who contributed greatly to the development and initiation of this project: Ernest Harrell, Ph. D., Richard Rogers, Ph. D., and Jeffrey Clark, Ph. D. Others who assisted greatly in the enlistment of subjects include Bruce Jones, Ph. D., of Neuropsychology Associates of Dallas; Ken Dobbins, Ph. D., of Wright Patterson Air Force Base Medical Center; Anthony Bird, Ph. D., of the Dayton VA Hospital; and Shirley Whitmer of the Ohio Brain injury Association. Thanks to Amy Moes for helping in the assessment of clinical subjects. I would also like to thank the many friends and family members who provided support and offered prayers on my behalf as I worked to finish this document. Lastly, I would like to thank my gracious wife, Diane E Isler, Ph.D., who understood exactly what it was like to struggle with dividing attention between work, family and school, and who assisted me in innumerable ways throughout the process.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The term malingering was originally used by the military to describe soldiers who complained of symptoms which allowed them to evade duty (Brussel & Hitch, 1943, cited in Nies & Sweet, 1994). Now the term is much more broadly applied, referring to accident neurosis, sham illness, simulation, dissimulation, and, among others, compensation neurosis (Gorman, 1982; Mendelson, 1985; Rogers, 1997). In each instance, the chance for secondary gain motivates persons to exagerate or even fabricate symptoms. For a variety of reasons, more people than ever are seeking compensation for alleged brain injuries; thus, neuropsychology is becoming increasingly important in the arena of forensic clinical assessment (Hall & Pritchard, 1997). For example, malingering may be used to excuse behavior since cognitive, emotional, and social functioning can all be influenced by changes in brain functioning. Also the direct potential for financial compensation and the indirect emotional and social rewards of faking make an appealing solution for many types of social, economic, and personal problems related to brain injury (Lezak, 1995). In addition, the emergence of managed health care and the increasing restrictions on resources necessitate that clinicians be able to adequately determine if patients are indeed suitable candidates for costly treatment. As a result, those patients who are seeking compensation and have a possible secondary gain for exhibiting neuropsychological deficits should be assessed for malingering.

The consequences of failing to detect malingering include unjustified monetary compensation, evasion of criminal prosecution, and acquisition of undeserved worker's compensation benefits (Franzen, Iverson, & McCracken, 1990), and the unavailability of certain limited resources for others who legitimately need them. On the other hand, repercussions for inaccurately classifying people as malingerers include withholding treatment from those who are sincerely in need. As a result, clinical interest has been recently heightened in the area of neuropsychology and malingering; 37 articles were published in this area between 1961 and 1990 (Franzen et al., 1990), while more than 50 articles were published between 1990 and 1995 (Haines & Norris, 1995). This intensity of research has provided specific methods to assess malingering in neuropsychological cases (Binder, 1992), although the utility of these measures has yet to be widely established.

Classification of Malingering

Malingering is defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association [APA], 1994) as "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives..." (p. 683). Four instances in which malingering should be suspected include: a) when a client presents for medical and legal reasons; b) when the client's reported symptoms are discrepant from the objective findings of an assessment; c) when the client does not cooperate with the assessment or with the treatment recommendations; and d) when the client is diagnosed with Antisocial Personality Disorder (APA, 1994). The detection of malingering is important for many legal and health professionals because of the cost in terms of litigation cases (Goebel, 1983). When

financial compensation may be awarded based on an assessment of the extent of injuries, clinicians must consider the motivational level of the client and the possibility of malingering.

The DSM criteria have been criticized for demanding that the clinician determine internal versus external motivation and for unfairly emphasizing the occurrence of malingering with antisocial personalities (see Rogers, 1990a, 1990b, 1997). In an effort to apply DSM-IV criteria as a threshold model (which determines when a clinician should thoroughly evaluate suspected malingering), Rogers (1990b) reported an unacceptably low sensitivity rate. Thus, an "adaptational model" was proposed in which malingering is viewed as an adaptive response to adverse circumstances (Rogers, 1990a, 1990b) . This model:

> ...assumes no link between sociopathy and malingering, nor, like the older pathogenic model, does it assume that malingered symptoms mask or emerge from genuine psychopathology. Instead, the adaptational model postulates that (1) malingerers perceive the assessment or treatment as adversarial, and (2) using a cost-benefit analysis, they consider malingering the best means of achieving their goals. (Rogers, 1997, p. 78)

Because of the inadequate empirical evidence to support any one diagnostic framework for malingering, the clinician must understand the limitations of a comprehensive application of the DSM-IV diagnostic framework (Ruff, Wylie, & Tennant, 1993).

Differential diagnosis. Malingering can be distinguished from Factitious Disorder by the motivation for the production of symptoms (APA, 1994). The classification of

Malingering is given when symptoms are manufactured in response to external incentives, whereas in Factitious Disorder, the internal incentives (e.g., the need to adopt a sick role) are the driving force behind the feigning of symptoms. Malingering also differs from Somatoform Disorders such as Conversion Disorder by the deliberate production of symptoms and the apparent external incentives connected with it.

Malingering can be difficult to diagnose, especially in the presence of authentic symptoms. For example, patients with mild to moderate damage may exaggerate their dysfunctions in order to appear worse than they actually are. In these cases, the clinician must make judgments about the degree to which symptoms are in fact causing impairment versus the degree to which external incentives are motivating the patient to present in a negative light. This picture may be particularly confusing with the high comorbidity of head injury and symptoms associated with depression and Post Traumatic Stress Disorder (PTSD; Pankratz & Binder, 1997).

At times, malingering is inappropriately suspected or diagnosed. The interpretation of neuropsychological test data is usually based on the assumption that the examinee has performed to the best of his or her ability (Golden, Zillmer, & Spiers, 1992), so clinicians should be aware of legitimate conditions which may impact an examinee to not give his or her best performance. Motivational variables which may compromise a client's test performance include level of arousal and cooperation (Golden et al., 1992). If a person has taken medications which cause a decrease in arousal, fatigue or drowsiness may ensue, thereby affecting the level of motivation and cooperation during testing. In this case, the data must be interpreted with caution if it is to be considered an appropriate

indicator of cognitive functioning. A clinician who is unaware of the situational factors which decrease motivation or level of arousal may misinterpret test data, thereby suspecting or misdiagnosing malingering.

Clients may become uncooperative with neuropsychological testing for several reasons, including fear of the testing procedure, concern over what the examiner thinks of them, inability to perform the tasks, and boredom. Clients often have misconceptions about the nature of neuropsychological evaluations, and may fear the examiner is going to use electrical equipment (e.g., that they will be "hooked up with wires" or "shocked"). Or they may fear the psychologist is measuring "craziness" (Golden et al., 1992). Most clients become more relaxed when an explanation of the procedure is provided and rapport is established. Clients may appear uncooperative when they fear the psychologist will make fun of them as deficits become apparent. Clients may also become uncooperative as the testing progresses either because of repeated failures or because of a lack of interest in the frequently repetitive tasks of neuropsychological assessment. Each of these factors should be taken into consideration when malingering is suspected and test performance is less than expected.

Prevalence. The true prevalence of malingering are unknown, in part due to undetected cases. Trueblood & Schmidt (1993) found an incidence of 7.5% in their sample of 106 consecutive admissions for neuropsychological evaluations. The 8 malingerers in their sample were identified by significantly below-chance performance on symptom validity testing. Estimates of faking of cognitive deficits range as high as 64% of personal injury cases (Heaton, Smith, Lehman, & Vogt, 1978) and 47% of workers'

compensation cases (Youngjohn, 1991, cited in Rogers, Harrell, & Liff, 1993). Estimates of malingering in patients seen in neuropsychology clinics have ranged from 33% to 60% (Greiffenstein, Baker, & Gola, 1994). The prevalence of malingering may be influenced by the setting or time at which the evaluation takes place (Rogers et al., 1993). For example, malingering may be common in cases involving litigation for monetary compensation following an injury, but less likely when compensation is not an issue.

Neuropsychological Outcome as a Result of Potential Financial Benefit

Some researchers (e.g., Bernard, 1990; Martin, Bolter, Todd, Gouvier, & Niccolls, 1993) have reported no connection between financial compensation and neuropsychological outcome. For example, Bernard (1990) randomly assigned participants to one of three conditions including control, malingering with a financial incentive and malingering without a financial incentive. Overall, the malingering groups did not differ significantly from each other in performance, although they did differ from controls. In fact, the malingering group provided with an incentive more often described themselves during a debriefing as being less confident they had succeeded in faking and in keeping the examiner blind to their attempts, while the participants asked to fake but not provided with an incentive were more confident about their attempts. Bernard (1990) hypothesized that cognitive dissonance theory might be used to explain these interesting results. In contrast to these findings, Binder and Willis (1991) reported that bona fide patients who were seeking compensation for minor head injuries performed more poorly than those who had a well-documented brain dysfunction but were not involved in litigation.

A recent study (Guilmette, Sparadeo, Whelihan, & Buongiorno, 1994) conducted with 50 patients referred for Social Security disability evaluations found a high incidence of malingering using a 36-item shortened version (Guilmette, Hart, & Giuliano, 1993) of the Hiscock and Hiscock Forced-Choice Procedure (Hiscock & Hiscock, 1989). Within the sample of Social Security disability claimants, 18% obtained scores of questionable validity (less than 90%, or 33, correct), while 20% obtained intermediate scores (between 33-35 correct). The intermediate scores were interpreted as evidencing at least some lack of motivation. In summary, as evidenced by their poor performance, approximately 1/5 of these disability claimants produced invalid and uninterpretable protocols (Guilmette et al., 1994).

In a separate study (Fee & Rutherford, 1988), the symptom rates of those seeking compensation at the time of the medico-legal evaluation did not differ significantly from the non-compensation group at 6 weeks post-injury. However, at the time of settlement there was a significant difference, with the compensation group having a symptom rate of at least two to three times higher than those not seeking compensation for injures. The litigation process seemed to have a deleterious effect on the symptom rate, and financial gain was described as an important factor in the production of symptoms (Fee & Rutherford, 1988). It is also possible that there is a group of people who are seeking compensation for injuries who are genuinely severely injured and have no incentive to fake. For example, someone who has MRI's or other significant evidence of injuries would have no reason to fake on neuropsychological measures. For this group it would that symptom reporting would be high but motivation to fake would be low.

Conflicting findings of performance level between simulation volunteers and patient groups may be attributable to the difference in the magnitude of incentives. Volunteers in many of these studies receive less actual or potential compensation than do patients. The typical incentive in simulation research is approximately $50.00 (e.g., Bernard, 1990) or less, while financial compensation in successful litigation may be in the millions of dollars. Thus, results of simulated studies must be interpreted with caution. For example, Martin et al., (1993) found that "motivational incentive" (p. 867) did not affect forced-choice performance, but the incentive, which was two dollars, may not have met the criteria for motivational incentive for the participants. Large monetary prizes for research participants are rarely offered because research budgets are restricted and because offering larger amounts of money as an incentive might be construed as coercion (Richard Rogers, personal communication November, 1995).

Neuropsychologists' Ability to Detect Malingering

Until recently, research of the simulation of brain damage on psychological tests has been meager, providing few concrete instruments for clinical use. Little guidance has been available for neuropsychologists endeavoring to detect malingering (Golden et al., 1992). In discussing the validity of measures for dissimulation, Rogers (1988) states, "Clinicians have been left, until recently, with a choice of either clinically relevant but methodologically questionable case studies or questionably relevant but methodologically sound psychometric and social-psychological studies" (p. 309). As a result of methodologically questionable or non-relevant measures, the examiner is often left with clinical intuition as the major defining process by which a diagnostic decision is reached

regarding malingering. The use of clinical intuition in the detection of malingering necessitates a broad awareness of the literature pertaining to neuropsychological sequelae of brain injury considering the heterogeneity of performance patterns (Brandt, 1988).

Some studies raise questions as to the clinicians' capacity to detect malingering in a reliable fashion. Rosenhan's (1973) provocative study of the detection of pseudopatients on a psychiatric ward yielded a 100% error rate. Several studies have shown that neuropsychologists also have difficulty correctly identifying malingering. For example, Heaton et al., (1978) found that 10 experienced neuropsychologists, when told that 50% of protocols examined were from fakers, were only able to achieve from chance to 20% above chance in the detection of simulators on the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), Halstead-Reitan Neuropsychological Test Battery (HRNB; Reitan & Davidson, 1974), and the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940). Poor detection rates were also noted in studies of pediatric and adolescent malingering (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988). Even when clinicians were informed that the protocols they were reviewing contained 50% malingerers and 50% actual cases, the rates of detection fell well below chance level.

Yet another study (Faust, Guilmette, Hart, Arkes, Fishburne, & Davey, 1988) found that training and experience of 600 neuropsychologists were not correlated with accuracy of clinical judgments regarding presence, localization, etiology, and prognosis of disorders. Participants were randomly selected members of both the National Register of Health Service in Psychology and Division 40 of the American Psychological Association.

Each participant was assigned case summaries which included drawings from the Aphasia Screening Test, and portions of the Halstead-Reitan Neuropsychological Test Battery (HRNB), Weschler Adult Intelligence Scale-Revised, and the Wechsler Memory Scale-Russell's Revision. Training and experience were not identified as factors in accuracy of diagnosis except for a greater tendency among more experienced clinicians to misdiagnose normal cases as abnormal. Methodological errors noted in this research include the poor selection procedures for respondents, the inability to address competency of skills in survey research, judgments made with test data which were not always the practitioners' preferred assessment approach, and the lack of ability to gather more relevant and appropriate information for each case (cf., Bigler, 1990; McCaffrey & Lynch, 1992). The Faust, Guilmette, et al. (1988) study has instigated new clinical research and increased the awareness of clinicians regarding the importance of clinical judgment and accuracy (Nies & Sweet, 1994).

Other factors which may contribute to neuropsychologists' inability to accurately detect malingering include the lack of definitive criteria for identifying malingerers and the unavailability of accurate feedback (Faust & Guilmette, 1990). Clinicians often do not know when they are wrong, so they have few chances to improve their skills of detection. The lack of reliable base rates of malingering within clinical populations also contributes to poor detection because clinicians do not know how often to expect deception (Faust & Guilmette, 1990).

In reviewing the recent literature reporting the lack of ability of neuropsychologists to detect malingering, McCaffrey and Lynch (1992) state, "the broad-sweeping

conclusions from these reports, while ignoring problems of internal and external validity, unjustly malign the field of clinical neuropsychology" (p. 245). In a more recent study (Trueblood & Binder, 1997), 60 psychologists reviewed test protocols from identified clinical malingerers and 27 reviewed the data of a documented head-injured patient. Error rates of psychologist judges in detecting malingering ranged from 0 to 25% across the cases of documented malingering. While these results are much better than those described by Faust et al. (1988), it is unclear whether this improvement in detection is a result of differences in methods between the studies, increased sensitivity and knowledge of this issue or improvement in malingering detection methods (Trueblood & Binder, 1997).

Clinicians' abilities to consistently detect malingering will likely continue to improve with advancements in the field of detection (Faust & Guillmette, 1990). New measures which are accurate and statistically sound are needed to enhance clinical judgment. An obvious method for clinicians to improve their ability to detect malingering is to become more familiar with the current literature (Franzen, et al., 1990; Haines & Norris, 1995).

General Strategies for the Detection of Malingering

Several strategies and measures have been developed for assessing malingering with regard to brain dysfunction. Perhaps the simplest strategy is an increased awareness of the possibility of malingering. Clinicians should be cautioned to consider malingering when a client could benefit from compensation for a disability or when a client's complaints exceed what the injury or illness would be anticipated to cause (Lezak, 1995).

The use of multiple measures and supportive evidence, such as prior test results and records which might reveal previously suspected malingering, will increase the likelihood of an accurate classification (Binder, 1993). Clinical interviews which include a careful history are imperative, and collateral interviews with spouses, relatives, friends co-workers, and employers are helpful in corroborating the information gleaned from the patient (Franzen et al., 1990). Although collateral information is useful, it is often difficult to quantify and thus compare to other groups.

Perhaps the most commonly used indicator of malingering is inconsistency in performance levels (Lezak, 1995). The clinician may examine the client's protocol for inconsistencies between subtests measuring similar attributes, or between scores on tests measuring sequential aspects of neuropsychological functioning. For example, attention/concentration skills are thought to be necessary prerequisites for the attainment of memory. In a recent study utilizing the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987) simulators tended to have lower attention/concentration scores than memory scores thereby exhibiting an inconsistency in performance levels (Mittenberg, Azrin, Millsaps, & Heilbronner, 1993). This method of performance evaluation, however, requires highly experienced and informed clinical judgment, and so may be influenced by individual interpretation. Several standardized dissimulation detection techniques and instruments have been developed to alleviate the variances which often result from inconsistently applied criteria. For example, simple memory tests are designed to be too easy for most people to fail; therefore, a high rate of failure strongly suggests faking.

Specific Strategies for the Detection of Malingering

Specific strategies for the detection of feigned neuropsychological deficits include: a) floor effect; b) symptom validity testing (SVT); c) performance curve; d) magnitude of error; e) atypical presentation; and f) psychological sequelae (Rogers et al., 1993). These strategies are described below.

Floor effect. The underlying principle of the floor effect is that an examinee who is attempting to fake will likely fail at a task that even most severely brain damaged persons can accomplish. Tasks which malingerers tend to fail while others do not may be included as integral parts of standard neuropsychological batteries. For example, the items of the Category test of the HRNB range in difficulty, and various items have been identified as rarely missed (Bolter, 1992; Bolter, Picano, & Zych, 1985). In a preliminary study of 50 brain damaged and 50 pseudoneurological patients, Bolter et al. (1985) identified 18 items missed by less than 5% of their non-malingering group. Using the identified items, Tenhula & Sweet (1996) were able to correctly classify 75.6% of their sample including controls, coached malingerers and brain damaged groups. However, Trueblood and Schmidt (1993) did not find that suspected malingerers made more errors than controls on the items identified by Bolter et al. (1985). In an attempt to replicate and extend his first study, Bolter (1992) identified 14 rarely missed items utilizing groups of 50 brain damaged and 50 healthy normal controls. Of 50 college controls, 98% missed none of these 14. Of 55 brain injured participants, 78% missed none and 16% missed one, with 98% earning a score of less than two errors. Poorer performance on these 14 simple items was interpreted as a pathognomic indicator of lack of motivation, attentional deficit, or an

attempt to overstate deficits. It was recommended that clinicians be suspicious of an examinee who misses more than two of these items since 98% of the participants with brain damage missed two or less (Bolter, 1992).

In an attempt to replicate and extend these findings, Tenhula and Sweet (1996) conducted studies to determine if the BCT could discriminate between groups of brain-injured, faking, and normal participants. Utilizing Bolter et al. (1985) 18 items, the overall hit rate was 84% with a sensitivity of 51% and a specificity of 98%. When comparing the 19 "Easy Items" derived from their study, the overall hit rate was 86% with a sensitivity of 57% and a specificity of 98%. Although the Easy Items of Tenhula and Sweet (1996) were somewhat more sensitive than those of Bolter et al. (1985), the difference does not appear to be appreciable. The best hit rate (92%) in this study was found for the number of errors on Subtests I and II. One large drawback to this study was its reliance upon comparison of the brain damaged group with undergraduate college students earning credit for their participation.

Other researchers have devised tests which are specifically designed to expose malingerers based on their tendencies to overestimate the degree of impairment, and therefore to perform more poorly than would be expected for someone with that particular disorder (Lezak, 1995). For example, Rey's 15-Item Memory Test (1964) is a technique designed to examine the possibility of dissimulation of memory complaints. The examinee is presented with a stimulus card on which 15 items are arranged in three columns and five rows. The card is exposed for 10 seconds, and then the examinee is asked to draw what is remembered. The 15 items on the stimulus card can be recalled by most people using only

three or four main chunks to remember the presented material (Lezak, 1995). The stimuli are simple in that items are redundant and easily chunked, but the measure superficially appears difficult. Typically, two recall scores are derived: 1) the total number of items drawn regardless of order; and 2) the number of correctly ordered sets of items. The addition of "spatial scoring" (correct within-row reproductions) has been shown in a recent study to improve the ability to predict group membership between litigating and non-litigating participants (Greiffenstein & Gola, 1996). Rey's technique may have gained support for the detection of malingering by many neuropsychologists because of its is ease of administration and scoring, but support for the utility of the measure to detect malingering is inconsistent.

Community participants who were asked to fake memory deficits on the Rey scored significantly lower than participants who were neurologically impaired, making more errors of omission (Paul, Franzer, Cohen, & Fremouw, 1992). Bernard (1990), however, found scores on the Rey did not distinguish between malingering and controls, but he conceded that placing the measure at the end of a battery of more difficult tests may have sensitized the participants to its "relative simplicity" (p. 71). In response to findings such as these, future research should examine the effects of test order and of simplicity in neuropsychological simulation studies (Bernard, 1990).

Although the Rey 15-item memory test has been shown in some studies to be effective with regard to its specificity, or its ability not to misclassify brain-injured patients, when used alone it appears to lack adequate sensitivity, or the ability to detect feigned impairment (Millis & Kler, 1995). In this study, only 57% of clinical malingerers

(identified by a below chance performance on a forced-choice memory measure) were correctly identified using a Rey cut-off score of 7, but none of the brain-injured subjects was misclassified. Other studies have also shown poor sensitivity ratings for the Rey, with values ranging from 14.5% (Schretlen, Brandt, Krafft, & Van Gorp, 1991) to 22.5% (Davidson, Suffield, Orenczuk, Nantau, & Mandel, 1991). Schretlen et al. (1991) found that using a cut-off of less than 9 items remembered identified less than 15% of the participants asked to fake mental disorders. Greiffenstein and Goal (1996) were able to improve specificity in their study by excluding "dense amnesiac" residing in supervised settings. Although the clinician or researcher can feel confident that false positives are infrequent with the Rey, true positives may not be as numerous as would be desired. There may also be settings for which the Rey's measures can not be meaningfully applied (e.g., Greiffenstein & Gola, 1996), and care should be taken when using the Rey 15-item memory tests even as a gross screening device for feigned memory loss (Rogers et al., 1993).

Symptom validity testing. Originally, symptom validity testing (SVT; Pankratz, 1979) referred to a technique in which the examinee was required to make 100 forced-choice decisions of a simple dichotomous problem (e.g., "Are these sounds the same or different?") involving the patient's complaint. Each of the two alternatives is presented 50% of the time. By chance alone, the binomial probability for random correct responses is 50%, so when an examinee's score falls below 50% correct, there is a likelihood of faking, as when malingerers "try too hard" (Rogers et al, 1993, p. 262) and miss more than chance. On the other hand, malingering can not be ruled out just because a

performance falls at or above chance levels. In a computerized SVT simulation study conducted by Martin et al. (1993), a majority of naive (21/39, 54%) and sophisticated fakers (36/40, 90%) attained scores at or above chance levels, suggesting that relying totally upon a binomial probability statistic for determining malingering creates the possibility of a high rate of false negatives. SVT has also been criticized for appearing too simple to the examinee. In this case the examinee may recognize that most people should be able to complete the task successfully and perform appropriately. Binder (1993) has extended the SVT to include a distraction technique in an attempt to build more face validity in the measure. Another limitation of SVT is the relatively small proportion of simulators that actually score below binomial probabilities (Rogers et al., 1993).

Although originally developed for sensory and perceptual complaints, SVT has also been applied to memory deficits (Lezak, 1995), and it has been expanded to include more sophisticated strategies, such as varying levels of item difficulty and using more than two alternatives (Rogers et al., 1993). Hiscock and Hiscock (1989) developed a forced-choice test applied to memory complaints, requiring examinees to identify which of two 5-digit numbers printed on a card are the same as a number seen just prior to a brief delay. The eight target numbers differ by two or more digits and are divided into three sets of 24 trials for a total of 72 trials. Each set has delays of 5, 10 and 15 seconds. Guilmette, Hart, and Giuliano (1993) determined that a cut-off at the 75% level (54 or fewer items correct) was appropriate to conclude malingering since non-patients in their study made no errors, and patients with brain damage and psychiatric inpatients made almost perfect scores. This study suggests that even a few errors on such an easy task should be viewed

with suspicion (Guilmette et al., 1993). In a separate study using the same method, Prigatano and Amin (1993) found that 6 suspected malingerers all performed at a level far below other groups (including unequivocal cerebral dysfunction, postconcussional syndrome and normals), but not significantly below chance, leading the researchers to state the Hiscock and Hiscock method may be useful in evaluating suspected malingers even when scores are not significantly below chance.

A variation on the Hiscock and Hiscock (1989) forced-choice test is the Portland Digit Recognition Test (PDRT; Binder, 1993; Binder & Willis, 1991), in which 72 trials of 5-digit numbers are presented auditorily at the rate of one digit per second. A distraction task is presented, requiring the participant to count backward aloud until interrupted with a visual recognition probe. Because of the distraction element, the PDRT is a test of working memory (Baddeley, 1976, cited in Lezak, 1995). Lezak (1995) cautions the clinician in the use of the PDRT for several reasons. First, the PDRT may be susceptible to finding poorer than expected performance in those with amnesia and frontal lobe injuries because amnestic patients perform poorly on distraction activities which are interposed between exposure to a stimulus and recall. Next, the PDRT's cut-off scores were originally derived from a group of brain-damaged patients in the VA system who were not seeking financial compensation for their injuries. Since many instances in which malingering is suspected involve possible compensation, the norms are of questionable value.

The Recognition Memory Test (RMT; Warrington, 1984) is an example of the application of SVT to an existing measure. The RMT is a forced, two-choice, two-subtest

(words and facial recognition) measure which provides a known chance level of correct performance (i.e., 50%). Millis (1992) found that 10 subjects with mild head trauma who were seeking financial compensation obtained significantly lower scores on both the word and facial recognition subtests of the RMT than did 20 subjects with a documented history of moderate to severe brain injuries who were not seeking compensation. A 90% specificity level and a sensitivity of 70% was obtained for those in the mild head trauma group, with scores less than 31 correct on the Word subtest. The lowest score on the Words subtest by a subject in the severe brain damaged group was 29; thus a score below 29 should raise a question of poor motivation on the part of a patient with only a mild injury.

Forced -choice testing (FCT; Pankratz, 1979) an early method of detecting fained symptoms assesses a specific ability (e.g. memory) by presenting a large number of items in a multiple-choice format and comparing the patient's performance to the likelihood of success based on chance alone. Although FCT certainly holds promise, Trueblood and Binder (1997) found that, contrary to their prediction, FCT data provided to psychologists reviewing test protocols did not enhance accuracy of identification. A significant relationship was found, however, between the availability of FCT results and confidence, with psychologists reporting more confidence in classifying malingerers when FCT scores were available. This indicates that clinicians may be able to rule out many other explanations when FCT data is available to them (Trueblood & Binder, 1997).

While much recent attention has been directed towards the development of detection procedures such as the symptom validity paradigm (e.g., Binder, 1990; Iverson

et al., 1991; Rawlings & Brooks, 1991), this type of detection strategy provides little additional diagnostic information. For example, if the patient is identified as a nonmalingerer, as many patients are, the test data adds little to the Neuropsychologist's ability to determine level of cognitive functioning. For this reason, measures are needed which are effectual for the evaluation of both malingering and cognitive functioning.

Performance curve. The performance curve strategy for detecting malingering is based on the supposition that malingerers will not consider varying levels of item difficulty when deciding which questions to fail (Rogers et al., 1993). Thus malingerers may be detected by comparing their performance curves to those of actual patients. The former would be expected to miss approximately equal numbers of easy and difficult items, while the latter would miss fewer easy and more difficult items.

The Dot Counting Test (DCT; Rey, 1941) identifies possible malingerers utilizing response latency to determine a performance curve (Lezak, 1995). The DCT assesses the validity of general or visual perceptual deficiencies by counting increasingly longer patterns of dots. Response time for those who are giving a good effort should increase as the number of dots increases, while those who are faking should show a different pattern. In this way the client's performance pattern is judged to be deviant and possibly malingering. The DCT was demonstrated to provide several scores that differed significantly between simulators and non-simulators (Binks, Gouvier, & Waters, 1997). In contrast, a study utilizing a computerized version of the PDRT (Binder, 1993) yielded longer average response latency scores for a head-injured patients compared to normal, coached, and uncoached malingerers (Rose, Hall, & Szalda-Petree, 1995).

Gudjonsson and Shackleton (1986) utilized the performance curve strategy in a simulation study with army personnel who were administered the Raven's Standard Progressive Matrices (SPM; Raven, 1977). Twenty-nine soldiers were asked to take the SPM under both honest and faking conditions. Their responses were compared with 27 soldiers of "modest intellect" and 25 soldiers evaluated for neuropsychological impairment. Participants' scores on Sets A and B were compared to Sets D and E using a formula for measuring linear trends or rate of decay (RD) across different levels of performance complexity, which was first described by Snedecor and Cochran (1967). A sensitivity rate of 82.8% and a specificity rate of 95.1% for the RD of the SPM were obtained using the following formula: (2A+B) - (D+2E) (Gudjonsson & Shackleton, 1986). The score for set C was not included in the analysis because of its central position among the subtests.

The performance curve method has also been applied to the Porch Index of Communicative Ability (Porch, Friden, & Porec, 1977, cited in Lezak, 1995), which has been used to test whether aphasia can be simulated. The simulated aphasia performance of 25 normal participants, divided into naive and well-informed groups were compared to a composite aphasia pattern. The simulators performed better on the more difficult items and had lower scores on the easier items than did the aphasics.

Frederick and Foster (1991) used a modified version of the Test of Nonverbal Intelligence (TONI; Brown, Sherbenou, & Johnson, 1982), creating a two-alternative, forced-choice, response technique. In a series of three studies, they examined performance curves, response consistency, and the product of the slope and consistency

ratio. By combining the three studies (N = 228 simulators, 14 patients, and 157 controls), and using a performance curve alone they were able to achieve a sensitivity of 26.3% with a specificity of 100% while the sensitivity for SVT alone was only 11.4%. Their measures of slope, score, and consistency were described as sensitive for only uninformed or obvious malingerers while a slope-consistency ratio product was found to be "highly sensitive to all types of malingerers but lack[ing] specificity with regard to high-scoring compliers" (Frederick & Foster, 1991, p. 600).

Magnitude of error. The magnitude of error strategy for detecting malingering consists of evaluating the qualitative or quantitative aspects of incorrect responses, which are not systematically considered in neuropsychological assessment (Rogers, et al., 1993). Absurd answers, or gross errors, are those which are very far from correct or very different from what is considered an appropriate answer. Although absurd or illogical responses may appear somewhat infrequently on neuropsychological measures, they may be important "pathognomic" signs of malingering (Trueblood, 1994). Approximate answers, or near misses, are responses which are very nearly correct. Bash and Alpert (1980) have suggested that malingerers may also be detected by their approximate answers.

Rawling and Brooks (1990), using a sample which consisted of persons who had sustained severe head injury as well as those with clinical evidence of simulation, devised the Simulation Index (SI), a method of objectively analyzing qualitative errors on the WAIS-R (Wechsler, 1981) and Wechsler Memory Scales (WMS; Wechsler, 1974). Fifteen types of errors were found to be made by simulators and five types of errors were

frequently made by head-injured participants. Objective scoring criteria for these 20 types of errors (e.g., "Primacy Errors," "Impossible Errors," and "Overtime") are included in the Simulation Index-Revised (SI-R; Rawling, 1993). A validation study of the SI-R with a large (n=388) mixed clinical sample including vascular dementia, schizophrenia, alcohol dependence, and brain-injured participants yielded false positives in over one-third of the patient protocols (Milanovich, Axelrod, & Millis, 1995). However, few false positives were found in the brain-injured sample, while higher rates were identified for those with vascular dementia, schizophrenia, and alcohol dependence. Although the SI-R may not be useful with non brain-injured populations, it is unique in its examination of the qualitative aspects of performance (Milanovich et al., 1995). In a separate qualitative analysis of WAIS-R responses, approximate answers, bizarre responses, and inconsistency in performance were found to be generally unhelpful for distinguishing between possible malingerers and controls (Trueblood, 1994). More systematic research is needed to devise methods for the evaluation of qualitative response errors.

Atypical presentation. Atypical presentation occurs when a patient misses more items or unusual items on a measure than would be expected in light of the symptoms reported. Atypical performance on neuropsychological measures may signify malingering (Lezak, 1995). An atypical presentation also occurs as variations in performance on tests measuring similar abilities (e.g., poor performance on CVLT and average performance on WMS-R) or on parallel-forms or readministrations of the same tests (Rogers et al., 1993).

Caution should be taken when considering the classification of malingering due to the atypical presentation of a patient, as inconsistency in the reporting of symptoms and

presentation have often been associated with functional disorders in patients with brain damage (Pankratz, 1988). Because of the lack of empirical data on atypical presentation as an indicator of malingering on neuropsychological tests, clinicians should be cautious when applying the method; gross errors, discrepancies, and inconsistencies should lead the examiner to further investigate the possibility of malingering with more empirically tested strategies (Rogers et al., 1993).

Psychological sequelae. Malingering may be detected by examining the frequency and intensity of reported psychological sequelae of minor or mild brain injury (Rogers et al., 1993), which include memory impairment, headache, dizziness, concentration difficulty, blurred vision, photophobia, ringing of the ears, irritability, fatigue, anxiety, and depression (World Health Organization, 1978; as cited in Youngjohn, Burrow, & Erdal, 1995). In order to distinguish between malingering and sincere deficits, the clinician should look for discrepancies between the client's presentation and common symptoms associated with neuropsychological impairments. However, the lack of systematic research regarding the common resulting symptoms of brain injury and their manifestation on neuropsychological tests limits the utility of this technique (Rogers et al., 1993). Distinguishing between pre- and post-injury symptoms may also be difficult, as some symptoms which patients with brain injury exhibit may be preexisting, while others are sequelae. Also, high base rates of neuropsychological symptoms have been reported in both personal injury litigants with no history of brain injury and in the normal, non-brain-damaged population (Lees-Haley & Brown, 1993; Gouvier, Uddo-Crane, & Brown, 1988).

Researchers have examined the level of knowledge of psychological sequelae following brain injury and found the knowledge in the general public to be limited. Aubrey, Dobbs and Rule (1986) found this was especially true with respect to the cognitive deficits and broad array of symptoms that may occur following a minor head injury. The lay people in their study emphasized physical complaints but showed a lack of "understanding or sympathy for memory problems, loss of concentration, and similar cognitive symptoms" (Aubrey, Dobbs, & Rule 1986, p. 845). Mensch and Woods (1986) found that the "lay conception of 'brain damage' involves primarily sensory-motor deficits in the form of exaggerated response times" (p. 63). Also, their lay subjects seemed to have little understanding of the lateralization of brain function (Mensch & Woods, 1986).

Research Designs for Malingering Studies

Rogers et al. (1993) describe two popular research paradigms for examining malingering: simulation and known-groups designs. Simulation research uses "normal" participants who are asked to fake neuropsychological impairments and are then compared to either other "normals" who were not instructed to fake or to a clinical comparison group such as brain-injury patients. The simulation design has both advantages and disadvantages. Since people who fake brain dysfunction are unlikely to admit they have attempted to fake (Cullum, Heaton, & Grant, 1991), simulation studies give researchers a means of identifying the response sets of people who are known to be faking. Also, simulation studies use experimental controls and a systematic comparison group (Rogers et al., 1993).

One of the difficulties with the simulation design is its assumption that the faking by normal participants is comparable to the performance of malingers (Cullum et al., 1991). In fact, the generalizability "to actual malingers in real-world settings" is unknown (Rogers et al., 1993, p. 257). A second problem with the simulation design is that the type of impairments faked by normals may differ both qualitatively and quantitatively from those presented by neurologically-impaired patients who have real deficits which can be easily exaggerated but not as easily detected by the clinician (Cullum et al., 1991). Third, the level of motivation to fake may affect the outcome. For example, many simulation studies provide evidence that lay persons have limited knowledge of neuropsychological assessment techniques and brain injury, but the simulators in most of these studies have been college students (Aubrey, Dobbs, & Rule, 1986) or other participants who have little interest or investment in their performance. This is not likely to be the case with malingerers who have been examined by several health care workers who have provided hints or information regarding the sequelae of brain injury through questioning of symptoms. In addition, published articles and books can be easily accessed by the well-motivated malingerer, providing a wealth of information regarding brain injury. There is some evidence that analog malingerers can accurately replicate self-reported postconcussive symptoms but are less able to simulate more objective measures (Martin, Hayes, & Gouvier, 1996). The knowledge of symptomology of brain injuries as well as information regarding neuropsychological assessment techniques should not be underestimated in the malingerer.

Known-groups design, which identifies known malingerers and compares them with actual patients on standardized measures, is not as frequently utilized, and is limited by the challenge of identifying and accurately categorizing malingerers (Rogers et al., 1993). However, the known-groups design has an advantage over the simulation design in its direct clinical applicability to genuine malingerers (Rogers et al., 1993).

A derivation of the known-groups design is the differential prevalence design, which assumes that the context under which the assessment takes place will create different rates of faking (Rogers et al., 1993). For example, patients seeking financial compensation for head injuries may be more motivated to malinger on neuropsychological measures than those not seeking compensation (Binder & Willis, 1991). The difficulty with the differential prevalence design is its inability to provide true prevalence rates and appropriate interpretations of group differences, which limit its utility for sound research (Rogers et al., 1993).

Rogers (1988) makes several specific recommendations for improving simulation research: 1) Participants who are asked to simulate malingering should be provided with incentives for success, such as a financial reward (which could be provided for eluding detection); 2) the instructions for simulation should be precise and emphasize the believability of the simulated responses; 3) participant compliance must be reviewed through manipulation checks to assess the accuracy of the involvement; and 4) analysis through discriminant functions and other multivariate techniques should be used to construct standardized indicators of dissimulation.

Battery Assessment Techniques in the Detection of Malingering

Formally developed neuropsychological batteries (e.g., HRNB) have been used in an effort to detect malingerers. Typically the goal of test batteries is to assess a wide array of the aspects of neurological function in order to more fully ascertain strengths and deficits in cognitive functioning. In the detection of malingering, inconsistency of pattern performance or differences between expected performance and disability have been utilized as markers of malingering (Lezak, 1995). In this way, each subtest in a battery can provide additional information to the clinician making a determination of malingering. No single test can adequately illuminate the intricacies of brain functioning (Reitan, 1966, cited in Goebel, 1983)

Halstead-Reitan Neuropsychological Test Battery. The Halstead-Reitan Neuropsychological Test Battery (HRNB) is comprised of at least 12 separate subtests which are described fully in Reitan and Wolfson (1985; 1993). The HRNB has been used for many years as a means of assessing neuropsychological deficits. The HRNB has no built-in validity measures but researchers have used it in an attempt to classify faking in simulation studies (e.g., Heaton et al., 1978; Goebel, 1983). The HRNB submeasures have also been utilized and proven moderately successful in the detection of faking (e.g., Charter, 1994).

Heaton et al. (1978) conducted a simulation study with the full HRNB. Responses of 16 patients with traumatic head injury were compared with those of 16 normals who were instructed to fake head-injury deficits. The malingering and head-injury groups evidenced similar overall impairment levels, but differed in their relative strengths and

weaknesses; the head-injury groups showed greater impairment on measures of abstract conceptual reasoning, cognitive shifting, and psychomotor problem-solving, while the malingering group demonstrated greater deficits in sensory-perceptual and motor functioning. Of the 37 measures examined with a stepwise multiple discriminant analysis, 30 correctly classified 100% of the participants. However, the degree of classification was likely inflated as a result of the number of predictor variables, which exceeded the total number of participants (Goebel, 1983).

Goebel (1983) used both subjective and objective analyses to determine whether patients with verified neurological impairment ($n = 52$) could be distinguished from non-impaired fakers and non-fakers $n = 202$ on the full HRNB. College student and community volunteers were assigned randomly to one of four faking groups (i.e., asked to fake right or left hemisphere or diffuse damage, or given nonspecific instructions) or to a control group. The subjective analysis, a blind clinical review of the protocols, yielded a classification rate of 94.4%. Discriminant function analysis revealed that the groups which were asked to fake lowered their level of performance somewhat, but in general were not successful in appearing impaired.

Luria-Nebraska Neuropsychological Battery. The Luria-Nebraska Neuropsychological Battery (LNNB; see Golden, Hammeke, & Purisch, 1980) has also been investigated to determine its utility in detecting faking. In a study utilizing a counterbalanced design in which psychologically naive subjects were asked to take the LNNB under honest conditions and then fake brain injury with the promise of a small financial reward if successful, those asked to fake brain damage were able to obtain scale

elevations indicative of significant deficits (Mensch & Woods, 1986). The researchers concluded that clinicians should compare the pattern of deficits obtained on the LNNB with known neurological conditions as a means of detecting faking. Although these types of suggestions to clinicians are somewhat useful in addressing the difficulty of diagnosing malingering, they add little practical help to the clinician beyond his or her own clinical judgment.

Detection of Malingering on Individual Neuropsychological Tests

In addition to neuropsychological test batteries, numerous existing individual measures have been examined for their utility in the detection of malingering. Clinicians frequently assemble their own test batteries which are comprised of portions of full batteries along with individual measures. Tests are then chosen or not chosen on a case by case basis depending on whether or not they assess the areas of functioning germane to a particular patient. Lezak (1995) and Smith (1975) are both examples of neuropsychologists who consider a flexible battery approach as providing the opportunity to add or subtract measures depending upon the needs of the assessment.

Wisconsin Card Sorting Test. The Wisconsin Card Sorting Test (WCST; Heaton, 1981) is a frequently used measure of neuropsychological functioning in which the patient is asked to sort colored cards according to an unknown shifting set of parameters while being given feedback on performance. When the WCST was administered to 62 brain-injured individuals, 34 psychiatric inpatients, 31 controls, and 58 normals who were instructed to malinger, significant differences between the malingerers and the other groups were detected on the number of categories completed, total number

of errors, total correct responses, and the number of nonperseverative errors (Knight, Webster, Goetsch, Malloy, & Greve, 1986, cited in Franzen et al., 1990). Sadly, no effort was made to determine the accuracy of the classification using the reported scores.

A separate study determined that the most typical pattern of groups consisting of brain-injured patients was that of low category scores with elevated perseverative errors, while a group of participants asked to malinger obtained lower mean scores on both number of categories and perseverative errors (Bernard, McGrath, & Houston, 1996). Yet some groups of true patients may show similar error patterns. For example, persons with severe general cerebral impairment, the elderly (6-10% of 75-80 years-old obtain Categories = 0), and those responding randomly may also produce similar patterns (Bernard et al., 1996). In light of these findings, the chances of a false positive seem rather high and malingering should be thoroughly corroborated by other means before a conclusion is reached.

Bender-Gestalt. The Bender-Gestalt (Bender, 1938) has nine designs which examine visuographic abilities. Since the Bender is quick and easy to administer, it is frequently used as a screening measure for neuropsychological deficits (Lezak, 1995). The Bender-Gestalt has gained some support for its utility in the detection of malingering (Bruhn & Reed, 1975). Employing scoring criteria which were developed in a pilot study, a clinician correctly classified 20 of 20 malingers (college students asked to fake) and 28 of 33 organic patients, yielding an overall classification rate of 91% (Bruhn & Reed, 1975).

Benton Visual Retention Test. The Benton Visual Retention Test (Sivan, 1992) is often utilized in neuropsychological assessment as a test of visuospatial memory. Benton and Spreen (1961) employed this measure in a simulation study comparing the performance of 47 college students and 23 medical patients asked to fake brain damage with 48 patients having a documented history of cerebral damage. The simulators scored significantly lower on this memory test than the brain-damaged patients, with more frequent distortion-type errors and fewer perseverations, omissions, and size errors than the clinical group. In a re-analysis of these data, a cutting score was developed which correctly classified only 65% of the simulators, while 27% of the brain-damaged were labeled malingerers and 35% of the simulators went undetected (Franzen et al., 1990).

Raven's Standard Progressive Matrices. The Raven's Standard Progressive Matrices (SPM; Raven, 1960; Raven, Court, and Raven, 1977) is a multiple-choice, paper-and-pencil measure developed in England which contains a series of visually-presented pattern-matching problems. Examinees must conceptualize numerical, spatial, and design features ranging from simple to very complex and abstract. Sixty items are grouped into 5 series which become progressively more complex and difficult. Two scores have been used with the SPM for detecting malingering: the rate of decay (Gudjonsson & Shackleton, 1986; described above) and the discrepancy score. The dependability of an examinee's scores across the sets can be analyzed by subtracting the statistically-derived expected score on each set (provided by Raven et al., 1977) from the examinee's actual score, thus creating a discrepancy score. If the examinee's score on one or more of the sets differs by more than 2, there is a possibility of decreased effort and the

"total score on the scale cannot be accepted at its face value as a consistent estimate of [the] general capacity for intellectual activity" (Raven et al., 1977, p.19). Gudjonsson and Shackleton (1986) found that the rate of decay score better classified fakers and nonfakers than did the discrepancy score, which yielded somewhat higher false positive rates in control and impaired groups.

California Verbal Learning Test. The California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987) is a frequently used measure of verbal memory which incorporates interference, free recall, and recognition into the testing for memory of a list of words. Trueblood (1994) found the CVLT to be sensitive to incomplete effort, with mildly head-injured patients, who were identified as giving poor effort based on symptom validity testing, differing from mildly head-injured control participants with respect to recognition hits and CVLT total. Other researchers (Millis, Putnam, Adams, & Rickter, 1995) have differentiated participants with moderate and severe brain injuries from those with mild head injuries who had been previously identified as giving incomplete effort by performing poorly on a forced-choice measure. All of the participants identified as providing incomplete effort were also in litigation and had external incentives for performing poorly. The portions of the CVLT which were identified as holding the most promise in the detection of malingering were the total for list A (trials 1-5), discriminability, recognition hits, and long-delay cued recall. The participants who were identified as providing incomplete effort exhibited two response styles, including a low number of correct recognition hits and a combination of a low number of recognition hits and a high rate of false-positive errors on the recognition trial following the 20-minute

delay. Using a linear discriminant function analysis, 91% of the participants were accurately classified (Millis et al., 1995).

Category, Speech-Sounds Perception, and Seashore Rhythm. The Category, Speech-Sounds Perception, and Seashore Rhythm tests (from the HRNB, Reitan & Wolfson, 1985) each yield scores which, if significantly different from random responding, suggest possible malingering. Random responding for these tests is best determined utilizing confidence intervals instead of specific scores (Charter, 1994). For example, using the method suggested by Brownlee (1965), a 95% confidence interval on the Category Test has a lower bound of 145 and an upper bound of 169. A score of 139 errors then is not considered reflective of random responding since it falls below the minimum cutoff score of 145 errors. Using data from 7 previous studies, Charter (1994) provides confidence intervals as well as norms for the Category, Speech-Sounds and Rhythm tests.

Posttest Debriefing in Simulation Research

Despite the call by leaders in the field to implement manipulation checks in simulation research (e.g., Rogers, 1988), few researchers do so (e.g. Binks et al., 1997; Mittenberg et al., 1993; Rawling & Brooks, 1990; Tenhula & Sweet, 1996). Posttest interviews or debriefing questionnaires provide a manipulation check which serves a myriad of valuable purposes, including the examination of participants' understanding of and compliance with instructions, since failure in either of these areas can greatly effect outcome (Rogers, 1997). Although participants are often assumed to have a clear understanding of the instructions and objective of simulation research, this may or may not

be the case. For example, Bernard (1990) reported that all participants in his simulation study were able to correctly restate the instructions they had received to fake, but none appeared to be aware of the true nature of the study. Most believed that the purpose of the examination was to determine if the examiners could be "misled" while a fewer number believed it was a memory test. Also, researchers cannot assume that even when faking instructions are understood, compliance is automatic. Goebel (1983) found that 10% of his faking group admitted that, contrary to their instructions, they had made no attempt to fake on any of the measures. In addition, 54% chose to fake only some of the tests, while only 36% made an effort to fake on every test. Most believed they were unsuccessful at faking the measures (Goebel, 1983). Debriefings can also help the researcher to gauge participants' motivation to succeed and follow directions. Bernard (1990) found that participants in all conditions (malingering, malingering non-incentive and control) reported trying at least moderately hard to comply with the instructions.

Debriefings may also be used to examine various strategies of simulation since participants may have very different ways of attempting to accomplish the task asked by the experimenter (Rogers, 1997). In summarizing the techniques used by his participants, Goebel (1983) noted that 30% gave wrong answers, 36% slowed their performance, 14% demonstrated motor incoordination, 2% attempted to show memory impairment, 2% ignored presented stimuli, 1.5% altered their emotional presentation, and 0.5% stuttered.

Different types of post-test debriefing strategies have been reported. Bernard (1990) used a questionnaire which included both a 5-point Likert scale and free response items. Goebel (1983) used an interview format. Rogers (1997) provides an excellent

overview of debriefing instructions which would lend itself nicely to creating a standardized debriefing questionnaire and should be reviewed by anyone planning research in the area of malingering. Regardless of the method used, more attention should be paid to the debriefing phase in malingering research (Rogers, 1997). By examining the types, frequencies, and effectiveness of different strategies utilized by participants, the researcher will be better able to competently develop more useful and well-standardized instruments for the detection of malingering.

Ethical Considerations in Malingering Research

Ethical considerations in the research of malingering have been discussed in terms of the coaching of participants in MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) simulation research (Ben-Porath, 1994; Berry, Lamb, Wetter, Baer, & Widiger, 1994), but the conclusions are immediately applicable to neuropsychological simulation studies. Coaching studies become ethically questionable since published research could be used by would-be malingerers as accessible training material (Ben-Porath, 1994), directly contradicting Ethical Standard 2.10 of the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct, which states:

> Psychologists make reasonable efforts to maintain the integrity and security of tests and other assessment techniques consistent with law, contractual obligations, and in a manner that permits compliance with the requirements of this Ethics Code. (American Psychological Association [APA], 1992, p. 1604)

In addition, published studies on malingering would have limited use for researchers or clinicians if specific strategies were not detailed.

Therefore, simulation researchers often experience ethical dilemmas in deciding what to publish based on the knowledge that sensitive information should be guarded and that clinicians or other researchers may benefit from their findings or attempt to replicate their studies. For example, although Frederick and Foster (1991) found that coaching was correlated with a decreased ability to detect malingering, they included specific instructions in their published article. In a separate article the authors acknowledge that the likelihood of coaching litigants for avoiding detection of malingering is increasing (Binks, Gouvier, & Waters, 1997) but go on to explain exactly how their participants were coached to avoid detection on a test of malingering. In publishing their findings, these researchers not only provided valuable information to assessors of persons suspected of malingering, but also made available information regarding proven techniques for faking.

The ethical dilemma for the researcher involves providing specific information regarding the procedures used in the study and the possibility that this published material could be misused (Ben-Porath, 1994). Some researchers have dealt with this predicament by "including only a brief synopsis of [the] coaching instructions in the article and ... releas[ing] the verbatim instructions only to those bound by the APA ethical mandate to protect the integrity of tests..." (Berry et al., 1994, p. 16). Others provide the instructions and cut-off scores to other qualified cinicians and researchers upon request (e.g., Tenhula & Sweet, 1996). In this manner, replications can be conducted while the ethical principles are maintained.

Purposes and Hypotheses of the Study

Purposes. Numerous techniques are available for detecting malingering, including: a) determining if a higher than expected number of easy items are missed (floor effect); b) determining if performance reflects poorer than chance responding (symptom validity testing); c) examining performance curves (proportion of easy to difficult items missed) since malingerers may not take into account differences in item difficulty; d) examining the number of gross errors (magnitude of error); and e) examining the discrepancy scores of the SPM. This study utilized well-established neuropsychological measures in an attempt to derive effective techniques for the detection of malingering while still providing meaningful diagnostic information for the clinician. The primary purpose of this study was to examine which of these techniques best differentiates participants asked to fake a brain injury from a group of participants with actual brain injury on the Raven's Standard Progressive Matrices (Raven, 1960) and Booklet Category Test (DeFilippis & McCampbell, 1979). Furthermore, exploratory analysis were conducted in order to assess differences regarding level of sophistication of knowledge of brain injury and level of motivation for secondary gain.

Hypotheses. The hypotheses of the study were based on specific techniques for detecting malingering. Because multiple-tasks simulation (e.g., not missing easy items and performance curve) may be more difficult to achieve (Rogers, 1997), this study was devised to examine multiple ways of detecting malingering on at least two separate measures which will provide the clinician with evidence to support or dispute a the

classification of malingering and/or data to support or disprove a diagnosis of brain dysfunction.

1) The first hypothesis examined the utility of the floor effect to detect malingering on the BCT and the SPM. A group comprised of participants from the general population who are asked to fake neurological impairment (NF) and a group of participants who were enrolled in or had completed a graduate course in neuropsychological assessment (SF) were expected to fail more of the obvious and easy items on BCT and SPM than a group of controls without history of brain injury (CN) or a group of patients with brain injuries not seeking compensation (BN) and a group of brain injured patients seeking compensation (BN).

2) The second set of hypotheses examined whether or not malingerers tend to perform poorer than chance. Due to their attempts to fake, NF and SF groups were predicted to miss more BCT items than would be expected by chance alone (derived from confidence intervals; Charter, 1994). In addition, due to their attempts to fake, NF and SF groups were predicted to miss more SPM items than would be expected by chance alone (derived from confidence intervals; (Charter, 1994).

3) The third hypothesis examined the efficacy of the performance curve strategy for detection of malingering on the SPM, which is especially suited for this strategy due to its progressively difficult nature. The differential rate of decay [(2A + B) - (D + 2E); Gudjonsson & Shackleton, 1986] provides a measure of performance curve for the SPM. NF and SF groups were expected to have a lower mean rate of decay than CN, BN, and BC groups.

4) The fourth set of hypotheses examined the efficacy of the magnitude of error strategy to detect malingering on SPM. NF and SF groups were expected to have higher numbers of gross errors (gross errors are defined as those answers which are farthest from correct and are not likely to be chosen by individuals giving their best effort) than BC, BN and CN groups. SPM Magnitude of Error was determined quantitatively by examining the responses of the BN and CN groups for those which were least frequently given.

5) The fifth hypothesis examined the efficacy of the SPM measure of discrepancy (Raven, 1977) to detect malingering. The NF and SF groups were expected to have higher mean numbers of discrepant scores on the SPM than BC, BN and CN groups.

Although no definitive hypotheses were made regarding the debriefing questionnaire, the responses of the faking groups were examined for typical ways of responding and knowledge of brain injury. Participants who reported an inability to comply with instructions to fake were excluded from the analysis (e.g., Frederick & Foster, 1991). Exploratory analyses were also conducted in order to assess differences regarding level of sophistication of knowledge of brain injury and level of motivation for secondary gain.

CHAPTER II

METHOD

Participants

Five groups of 30 participants each were included in this study. Clinical patients with brain injury were placed in one of two groups: those who were involved in litigation seeking compensation for their injuries at the time of testing comprised the brain-injury compensation (BC) group, and those who were not involved in litigation or had never sought compensation comprised the brain-injury non-compensation (BN) group. Two groups consisted of non-brain-injured persons who were instructed to fake brain injury, one sophisticated with respect to knowledge of brain functioning (SF), and the other naive with respect to brain functioning (NF). A comparison group comprised of individuals who were naive with respect to brain functioning (CN) was also included in the analysis.

Data for the BC group were collected at a small private practice in Dallas, TX by two full-time neuropsychologists who supervised two part-time testing technicians (the author and another UNT graduate student), and who used the HRNB approach in diagnosis and treatment planning. Each participant in the BC group received additional measures which varied according to deficits evaluated. Many of the people who initially presented to the clinic for neuropsychological evaluation were in litigation as a result of their injuries, so participants for the BN group were solicited from various sources, including: two group homes in Dallas, TX; the neuropsychology clinic at the Wright-

Patterson Air Force Base Medical Center in Dayton, OH; the Dayton Veterans Administration Hospital; and brain-injury support groups in the greater Dayton area. Patients in these groups were typically diagnosed with traumatic brain injury, post-concussional disorder, cerebrovascular accidents, infectious diseases, and cerebral anoxia. The participants were selected on the basis of a precipitating brain injury and then classified into groups based on their involvement in litigation for compensation.

The SF group was comprised of psychology graduate students who were enrolled in or had completed neuropsychological assessment training, and who were paid $15.00 for their participation. Participants in this group were asked to conform their responses on two neuropsychological measures to be consistent with those of a person with brain injury who was seeking compensation.

In an effort to diversify the other comparison groups, remaining participants were solicited from advertisements placed around the community and they were paid $15.00 for their participation. The NF group was comprised of persons without brain injury or training in neuropsychological assessment who were given the same instructions as the SF participants. The CN group included participants who were instructed to give their best effort on all measures. CN participants were considered naive with respect to brain function and head injury, but this could not be absolutely guaranteed.

Instruments

Screening measure. The Adult Neuropsychological Questionnaire (ANQ; Melendez, 1978) was used as a screening measure as well as to assess whether participant groups differed with regard to symptom pattern. The ANQ groups 54 questions under the

following headings: general health, substance abuse, psychiatric problems, general neurological, right hemisphere, left hemisphere, subcortical cerebellar, spinal, and sensory/perceptual. ANQ questions include commonly reported symptoms for head-injury patients such as, "Do your hands tremble sometimes?" and, "Have you hit your head lately?" Although often administered in an interview format, the ANQ can be used as a checklist which participants complete on their own.

Intellectual measures. Estimates of intellectual functioning were derived either from WAIS-R scores or from Slosson Intelligence Test (SIT; Jensen & Armstrong, 1985; Slosson, 1982) scores. Many of the clinical participants were administered the WAIS-R as part of their full neuropsychological evaluations. The SIT was used with the remaining participants because a) it is a relatively brief instrument and b) the lack of exposure by graduate students to this measure made it the best choice to use with the Sophisticated Fakers (graduate students in psychology) who were known to be familiar with most other brief intelligence tests.

SIT item types are adapted from the Binet and Gesell Developmental Schedules (Kaufman, 1990) and include vocabulary, knowledge of facts, arithmetical reasoning, and short-term memory. Correlations between SIT and WAIS-R IQs of 0.83 (Verbal), 0.51 (Performance), and 0.78 (Full Scale) have been obtained (Klett, Watson, & Hoffman, 1986). The higher correlation with Verbal IQ is expected because of the heavy verbal loading of the SIT (Kaufman, 1990). Ordinarily, the SIT should be used only as a screening instrument because a) it is based on questions from the original Binet, which has gone through many revisions; b) it is quick and easy for clinicians to administer; and c) it

is based on a narrow and unrepresentative sample (Kaufman, 1990). Despite these drawbacks, the use of the SIT in this study was necessitated by the inclusion of participants who were knowledgeable about other intellectual measures.

Raven's Standard Progressive Matrices. Raven's Standard Progressive Matrices (SPM), initially published in 1938, is reported to a) measure a person's capacity to observe relationships between meaningless figures by developing a systematic method of reasoning, and b) assess the examinee's ability to conceptualize spatial designs and numerical relationships (Lezak, 1995). Although the SPM does not discriminate well between undifferentiated groups of patients with right and left hemisphere damage, it is a frequently used measure of damage with organically impaired patients (Lezak, 1995). Factor analytic studies indicate that SPM measures Spearman's 'g' or general intellectual functioning, and has small loadings on visuo-spatial skills (Green & Kluever, 1992; Raven et al., 1977).

The SPM consists of 60 items which are divided into 5 sets of 12 figures in a booklet form. Each set begins with an item which should be self-evident with one part removed. There are from six to eight choices pictured for each item, with only one to correctly complete the pattern of the figure. The examinee's score (total number of correct answers) is converted to a percentile and provides an index of intellectual capacity. The items become progressively more complex. The first set (A) consists of incomplete figures with the missing part as one of the six choices given below the figure, so that the participant's task is one of simple pattern matching. Later sets are characterized by

increasing complexity requiring reasoning by analogy and mathematical concept formation (Lezak, 1995).

Raven et al. (1977) report generally good reliability in terms of internal consistency and retest reliability. Most of the studies providing consistency data yield correlations of at least $r = .90$. The reliability of the measure seems to decrease with younger ages (e.g., $r = .60$ for children under 6 years of age; Raven et al., 1977). Test-retest reliability rates vary from approximately $r = .90$ for short-term studies to $r = .80$ for longer intervals (Raven et al., 1977). Retest studies of up to one year later have also found acceptable retest reliability (Raven et al., 1977). Scores on the SPM tend to reach their highest point at the age of 14, remain relatively constant for the next 10 years and thereafter begin to decline slowly (Raven et al., 1977).

Validity measures of the SPM have been less predictable, with the concurrent and predictive validities varying with the age, sex and homogeneity of the sample (Raven et al., 1977). For adolescents and children who speak English, correlations between the SPM and Binet and Wechsler scales range from $r = .54$ to $r = .86$. The SPM correlates more highly with nonverbal and performance scales than with verbal intelligence tests. Similar results have been reported for adults, with correlations between $r = .75$ and $r = .88$ for SPM and WAIS scores. These findings do not hold cross-culturally, even though Raven et al. (1977) initially postulated that the test would be relatively free of culture bias. While SPM apparently requires no language or academic skills for success, there are educational influences on test performance (Lezak, 1995) which likely contribute to the cross-cultural differences. No gender differences on the SPM have been found. In terms of predictive

validity, correlations ranging up to r = .70 have been found between the SPM and measures of scholastic achievement (Raven et al., 1977). However, more research is needed to determine why predictive validity appears strongest for those scoring in the upper ranges as opposed to those in the lower ranges (Raven et al., 1977).

Halstead Category Test/Booklet Category Test. Halstead (1947) selected seven measures, including the Category Test, based on their ability to differentiate patients with frontal neuropathology from either non-frontally involved patients or normal controls. The Halstead Category Test (HCT) initially had 336 visually-presented stimulus figures for which the examinee was asked to abstract the underlying organizing principle for each item. Reitan and Davison (1974) later adapted and included this measure of "concept formation and the ability to apply organizing principles in performance of complex procedures" (p. 58) in the HRNB. The HCT requires complex concept acquisition and assesses visual memory, visuospatial reasoning, and the capability to translate visual information into verbal output (Rothke, 1986). Although it is one of the best overall measures of cerebral dysfunction in the HRNB, the HCT is not considered capable of localizing or lateralizing neuropsychological function (Reitan & Davison, 1974; Reitan & Wolfson, 1985).

The original version of the HCT, still utilized by many neuropsychologists, is mechanized and somewhat cumbersome (Lezak, 1995). It rewards the correct answer with a chime while error responses are recognized by a buzz, which has been considered by some patients and examiners as aversive. Several different versions of the category test are available, including the Booklet Category Test (DeFilippis & McCampbell, 1979) and

computerized versions (e.g., Beaumont, 1975). A study examining these different forms of the Category Test concluded there were no significant differences between three separate versions (Mercer, 1994).

Booklet Category Test. The Booklet Category Test (BCT), developed by DeFilippis and McCampbell (1979), will be utilized in this study. The BCT contains 208 visually-presented items arranged in seven subtests. The first six sets of items are organized around a principle for solving them, while the seventh is made up of previously viewed items. The examinee is required to discern the principle for solving each of the individual subtests. The examiner gives no real information regarding the principle, but the examinees are given feedback as they are told each choice is either "correct" or "incorrect."

The first set begins with roman numerals and the participant is required to point to a number between one and four. The geometric figures (in this case roman numerals) should remind the examinee of a number between one and four. Further examples include the third set which is organized on the principle of differences. The examinee is required to determine which one of the geometric figures differs from the rest and point to the number relative to its position. The fifth set displays geometric figures constructed of solid and dotted lines. The examinee is required to discern the principle of proportion. In this case the number of solid lines is the correct answer. The seventh set contains a memory component and tests the examinee's recall. In each subtest, the examinee is required to figure out the principle for solving the test and respond by pointing. The total number of errors is recorded as a summary score. The age of 40 years appears to be a

turning point in scores on the HCT; beyond 40 error scores begin to climb, with a rapid rise after age 60 (Lezak, 1995).

The BCT is a more easily administered test than the mechanized (HRNB) version (Lezak, 1995). The administration time is shorter because the examiner is able to flip quickly through cards when the examinee solves them easily. Items can also be discontinued when a set appears to be frustrating to the examinee (e.g., repeated failures).

Neuropsychological Simulation Debriefing Questionnaire. The Neuropsychological Simulation Debriefing Questionnaire (NSDQ) was developed for use in this study by the researcher (see Appendix A). The NSDQ consists of 13 questions, two of which are open-ended questions that examine the participants' comprehension and recall of the research instructions. The remaining 11 questions are multiple choice and evaluate the participants' self-report of effort, compliance, knowledge of brain function, means utilized in faking, and level of perceived success.

Procedure

Recruitment of participants. NF and CN participants were recruited via advertisements placed in workplace break rooms, apartment complex laundry and mail rooms, and community bulletin boards. Business card-sized advertisements were also distributed. Advertisements included a brief description of the study, the researcher's phone number and university affiliation, and mention of the monetary incentive ($15.00) for participation. Community volunteers responding to advertisements were given a brief summary of the study along with available testing times.

SF participants were recruited from graduate-level neuropsychological assessment courses and pre-doctoral internship sites. They were offered $15.00 for their participation.

BC participants were recruited from patients presenting to a private practice in Dallas, TX for neuropsychological evaluation as part of the pre-trial litigation process. Each patient was given a brief description of the study and was asked for permission to include their test data in the study. No monetary compensation for participation was offered.

BN participants were recruited from group homes, support groups, and patients presenting to the Dayton VA or Wright Patterson Medical Center for neuropsychological assessment. In some cases, names were furnished by providers and potential participants were contacted by phone, given a brief description of the study, and asked to participate. In other cases, the researcher visited brain-injury support groups and recruited subjects. Those who were receiving neuropsychological assessments not as a part of litigation were given a brief description of the study and asked permission to use their test data. No monetary compensation for participation was offered to BN participants.

All participants were assigned code numbers to ensure anonymity and confidentiality. A master list was kept of the participants in the NF, SF, and CN groups, which was used to distribute the incentive money to participants who were able to avoid detection.

Screening and testing of nonclinical participants. Nonclinical (NF, SF, and CN) participants took part in a single testing session, which lasted approximately 1½ to 2

hours. Each nonclinical participant was tested by the researcher or by an assistant with a Ph.D. in clinical psychology. First, a brief description of the study was given and each was asked to review a written description of the study and sign the informed consent form (see Appendix B). Demographic information, including gender, handedness, age, socioeconomic status (SES), ethnicity, and educational level was collected. Next, the ANQ was administered to screen for common symptoms of brain injury. Nonclinical participants who presented with definitive neurological symptomology, psychiatric symptoms, or current substance abuse were excluded from the study. Each participant in these three groups was then administered the SIT as a brief screening measure of intelligence, with the instructions to give their best effort on the test.

Next, the community volunteers were asked to choose a slip of paper out of a box and read it without revealing its contents to the examiner. The box held 60 slips of paper of the same size, 30 of which contained instructions for faking and 30 of which contained instructions for giving their best effort. In this way, assignment to the NF or CN group was made while the examiner remained blind to condition. Once participants indicated an adequate understanding of the instructions, the BCT and SPM were administered.

> The written scripted scenario given to NF and SF participants was as follows: You are participating in a series of neuropsychological tests as part of a research project. Please attempt to fake these tests. Your part in this study is to take these measures while playing the role of a person with brain damage. You were a passenger in an automobile involved in an accident. The driver of the other vehicle was at fault and you are entitled to have your medical costs paid for. While taking

these tests, you should alter your performance to respond the way you think a person with brain damage would respond. Please do not be obvious as this could result in losing the settlement and in severe court penalties. I will not be able to coach you or offer you any other suggestions as to the best way to do this. You are free to change your responding in any way you think would produce the most believable and realistic results. After the results of this study have been analyzed, if you are able to avoid detection in your responses, you will be eligible for a drawing of $50.00. If you are selected, you will be mailed a check. Do you understand these instructions? (adapted from Frederick & Foster, 1991; Goebel, 1983)

The instructions given to CN participants were as follows:

Please give your best effort on these tests. Take these measures as you normally would, giving your best and true effort. In addition to your $15.00 for participating in this study, your name will be entered in a drawing for $50.00. If your name is picked for the $50.00, you will be mailed a check. Thank you for your participation.

Following the completion of the ANQ, Slosson, BCT, and SPM, each of the SF and NF participants was administered the NSDQ in order to ascertain the strategies and/or knowledge the participants had concerning neuropsychological deficits and faking.

Testing of clinical participants. BC participants were tested at Neuropsychology Associates of Dallas, a private practice in Dallas, Texas. All screenings and neuropsychological evaluations were administered by the researcher or another UNT

graduate student, both of whom were employed as neuropsychological testing technicians. Prior to testing, patients were given a description of the study and asked to sign an informed consent form if they agreed to have their data included in the study. BC participants were then administered the full HRNB, including the BCT, as well as the SPM WAIS-R, and Rey. BC participants were directed individually by the examiners to do their best during the entire examination, receiving the following instructions along with the standardized instructions for each measure:

> "You are participating in a series of neuropsychological tests. Please give each test your best effort."

From their test data, demographic variables, WAIS-R, ANQ, BCT, SPM, and Rey scores were extracted for use in this study. Identifying information was deleted from the records and code numbers were assigned to assure anonymity.

BN participants were tested in their homes by the researcher or an assistant with a Ph.D. in clinical psychology. They were not administered an entire neuropsychological battery, but received only the core battery of this research study which included demographic data as well as the ANQ, SIT, BCT, SPM, and Rey. BN participants were directed individually by the examiners to do their best during the entire examination, receiving the following instructions along with the standardized instructions for each measure:

> "You are participating in a series of neuropsychological tests. Please give each test your best effort."

CHAPTER III

RESULTS

Ethical Consideration of Results

Publishing the results of malingering research could potentially lead to the coaching of strategies derived from the research (Ben-Porath, 1994; Berry et al., 1994). As a result, the items used in the variables calculated for this research will not be reported here, but will be made available from the author upon request.

Sample Characteristics

Homogeneity between the groups was determined with respect to demographic variables. No differences were found for handedness, $\underline{X}^2$ (4, $\underline{N}$ = 150) = 5.042, $\underline{p}$ = .283, participant income, $\underline{X}^2$ (24, $\underline{N}$ = 150) = 23.612, $\underline{p}$ = .484, or ethnicity $\underline{X}^2$ (16, $\underline{N}$ = 150) = 12.008, $\underline{p}$ = .743 (see Table 1). Separate chi-squares were run for Caucasian versus Minorities and income collapsed in $20,000 segments (e.g., 0-20K, 20-40K); these were also non-significant. Right-handers comprised 91% of the sample while left-handers accounted for the remaining 9 %. The largest percentage of the sample (29%) had annual incomes of $10,000 or less, while 11% of the sample had incomes in excess of $50,000. The majority of the sample (90%) were Caucasian with Hispanics, African Americans, and Asians comprising 5% or less each. Differences among the groups were found for gender, $\underline{X}^2$ (4, $\underline{N}$ = 150) = 16.885, $\underline{p}$ = .002 (see Table 1); and for age, $\underline{F}$ (4, 145) = 6.169, $\underline{p}$ <

.001 (see Table 2), mean Adult Neuropsychological Questionnaire (ANQ) score, F (4, 145) = 52.021, $p < .001$ (see Table 3), and IQ, F (4, 145) = 25.103, $p < .001$ (see Table 4).

The average age of the participants in this study was 36.3 years. The SF group was the youngest with a mean age of 29.3 years, while the NF, CN and BN groups fell in the middle (respective means of 32.8, 37.1, and 39.6 years). The BC group contained the oldest individuals with a mean age of 42.5 years. Both the NF and SF groups were significantly younger than the BC group ($p = .038$ for NF and $p = .001$ for SF). The mean ages of the CN and BN groups did not differ significantly from any other groups.

Of the 150 participants, 67 were male and 83 were female. The BC group contained an almost even ratio with 52% males and 47% females. The BN group contained 73% males and 27% females, the opposite of the remaining three groups (CN, NF and SF) which contained ratios of approximately 30% males and 70% females each.

With respect to the ANQ scores, the clinical groups did not differ from one another (mean of 20.7 for BC and mean of 16.73 for BN), but their scores were significantly higher than any of the nonclinical groups. Likewise, the nonclinical groups did not differ from one another with respect to ANQ scores. The SF group had the lowest scores, with a mean of 3.10 and a range from 0 to 7.

The average IQ score for this study was 103 across groups. As expected, the SF group, which was comprised of graduate students, had significantly higher IQ scores than any other group, with a mean of 122.17. The clinical groups (BN and CN) did not differ

from one another with respect to IQ. The NF group (mean of 102.63) was found to differ from the BC group (mean of 91.73) but not the BN group (mean of 94.37).

Comparisons between the BC and BN groups on clinical variables yielded significant differences with respect to diagnosis, X^2 (5, N = 60) = 11.106, p = .049 (see Table 5); months post injury, t (58) = 3.565, p = .001; and number correct on Rey Memory for 15 Items, t (58) = 2.240, p = .029 (see Table 6). Overall, the BN group contained a more varied diagnostic pattern. Fifty percent of the BC group was comprised of closed head injury (CHI) patients while 83 % of the BN group had a diagnosis of CHI. The BN group also contained 3 post-concussive syndrome (PCS) and 2 patients diagnosed with anoxia. The BC group contained diagnoses of 8 PCS, 2 vascular accidents (CVA), 1 anoxia, 3 toxic exposures and 1 electrical shock. The mean months since injury for the BC group was 34.30 while the mean for the BN group was 93.97. The mean Rey score for the BC group was 13.03 while the mean for the BN group was 11.60.

Total BCT (TBCT) and SPM (TSPM) scores were also different between groups (see Table 7). A One-way ANOVA revealed significant differences between the groups with respect to TBCT (F (4, 145) = 12.012, p < .001). The NF group produced the highest mean errors at 94 and the CN the least at 42. The means for both the BC and BN groups were comparable at 69 and 67 respectively. Utilizing the Tukey HSD for post hoc analyses (see Table 8), the NF group mean was found to be significantly higher than BC (p = .009), BN (p = .003), and CN (p < .001). The SF group differed significantly only from the CN group (p < .001). The NF and SF groups were not significantly different from one

another (p = .104). Thus, Tukey HSD yielded three homogeneous subsets: a) CN; b) BN, BC, and SF; and c) SF and NF.

A One-way ANOVA revealed significant differences between the groups with respect to TSPM as well (F (4, 144) = 31.442, p < .001, see Table 7). The mean score for the CN group was the highest at 52 while the NF group was the lowest at 23. The BC and BN groups were essentially commensurate with means of 41 and 39 respectively. Since the group variances were not homogeneous (Levene Statistic = 7.5, p <.001), the Tamhane T2 was utilized in the post-hoc analyses[1] (see Table 8). The NF group mean was significantly less than all other groups (BC, p < .001; BN, p < .001; CN, p < .001; and SF, p = .045). The SF group mean was significantly less than the means for BC (p =.013) and CN (p < .001).

Normality of the Distributions of Variables

The normality of the underlying distributions for each comparison variable by group was examined using the Shapiro-Wilk test for sample sizes less than 50. Normality was rejected at $p \leq .01$, and all continuous comparison variables except SPMRD (Standard Progressive Matrices Rate of Decay) were found to be non-normally distributed (see Table 9). General conformity to the normal curve and outlying variables were also determined for each variable by group using boxplots and stem-and-leaf diagrams to check for outliers. Exploration of the data with boxplots indicated that the within-group distribution for the variables was asymmetric and the variances were not equal across

---

[1]Tamhane's T2 is a conservative pairwise comparisons measure which is based on the t-test and is appropriate when variances are unequal.

groups. Levene's test for homogeneity of variance revealed unequal variances for all comparison variables except total BCT score (TBCT) and SPMRD (see Table 10). No transformation of the data was able to symmetrize the distributions or linearize the relations among variables. This was likely due to the way in which the variables were derived. Each variable was chosen because of an a priori hypothesis regarding its ability to distinguish between groups of fakers and non-fakers. Cut-offs were generated which automatically placed the distributions for groups at opposite ends such that the data was skewed highly left for some groups and highly right for others, thus making any type of transformation ineffective.

Correlations Between Demographic and Comparison Variables

The four demographic variables which differed significantly among the groups (Age, ANQ, IQ, and Gender) were correlated with the comparison variables to determine if covariates were required for between-groups analyses. The Pearson correlation coefficient was used to correlate continuous variables, Eta was used as a measure of association when one variable was nominal and the other continuous, and the contingency coefficient was used with nominal by nominal data.

Significant Pearson correlation coefficients were identified between two demographic variables (ANQ and IQ) and several comparison variables (see Table 11). ANQ was significantly negatively correlated with SPMGE ($r = -.222$, $p = .006$), and positively correlated with SPMRD ($r = .267$, $p = .001$). Intelligence was found to be significantly negatively correlated with TBCT ($r = -.202$, $p = .013$) and SPMRD

(r = -.324, p = .000). Neither age nor gender was significantly correlated with any comparison variable.

Analysis of Hypotheses

Hypothesis 1. The first hypothesis examined the utility of the floor effect, operationalized as validity indexes, for detecting faking. The Isler Validity Index (BCTIVI) was empirically derived for BCT by examining a frequency table of BCT items. The items on subtests III-VII which were missed by less than 5% of the groups BN and CN were included in this index (see Table 12). Since all but one of the 28 items on subtests I and II were missed by less than 5% of the sample, none was included. The BCTIVI was comprised of 13 items: 2 from subtest V, 7 from subtest VI, and 4 from subtest VII. BCTIVI items were somewhat different from those determined by Bolter et al., (1992) and Tenhula and Sweet (1996). The 6 items which were common to all three studies, labeled Combined VI (BCTCVI), included two items each from subtests V, VI, and VII. Thirteen items were common to at least two out of the three studies (BCT Combined Plus VI, or BCTCPVI), including 3 from subtest V, 7 from subtest VI, and 3 from subtest VII. The items derived by Bolter et al. (1992; BCTBVI) and Tenhula and Sweet (1996; BCTTSVI) were also included in the analysis of this hypothesis. A validity index for SPM (SPMVI) was derived by examining the item frequencies for groups BN and CN and included those items missed by less than 5% (see Table 13).

Since the data for the VI's violated the assumption of normality of distribution, the Kruskal-Wallis, a nonparametric version of the one-way analysis of variance for independent samples, was utilized to test for difference among group means. Significant

differences were found for all the VI's at the $p < .001$ level (see Table 14). Further analyses were conducted using parametric tests in an attempt to identify where group differences lay. Because the VI's were highly intercorrelated ($p < .001$, see Table 15), a MANOVA was used to ensure that actual differences were not masked by the significant intercorrelation of the scores. A significant main effect for group was found, Pillai's F $(20, 576) = 3.77$, $p < .001$ (see Table 16). Because the results of the Kruskal-Wallis and MANOVA tests were commensurate, the parametric measure was considered to be robust with respect to violations of its assumptions of normality of distribution.

Planned comparisons tests supported the hypothesis that faking groups scored significantly higher on these validity indexes than did non-faking groups ($p < .001$ for all comparisons, see Table 17). Post-hoc analyses were also conducted to further illuminate where groups differences lay (see Table 18). Since the groups were found to have unequal variances, Tamhane's T2 was used. The NF and SF group means for BCTBVI were significantly higher than those for BC ($p = .002$ and .026, respectively) and CN ($p =$ .001 and .002, respectively; see Table 17). On the BCTTSVI, NF had a significantly higher mean than BC and CN ($p < .001$ for both) and BN ($p = .041$). SF was significantly higher than both BC and CN ($p = .011$ and $p < .001$, respectively), but was not different from NF ($p = .232$). The BCTIVI showed similar significantly higher mean scores with NF being higher than BC and CN ($p < .001$ for each) and BN ($p = .048$). SF was also significantly higher than BC and CN ($p = .012$), but not BN ($p = .37$). Utilizing the BCTCVI, the NF groups means were also found to be higher than BC and CN ($p = .005$); however, SF was significantly higher than only CN ($p = .014$). For BCTCPVI, NF and SF

group mean scores were significantly higher than both BC and CN ($p = .009$). On SPMVI, the NF group mean was significantly higher than all other groups (BC, BN, CN, SF; $p < .001$) and the SF mean was greater than CN ($p = .002$) and less than NF ($p < .001$).

Hypothesis 2. The next hypothesis examined whether or not the faking groups (NF, SF) would miss more items than would be expected by chance on either the BCT or SPM. BCTChance was calculated for the BCT as any score greater than 145 (95% confidence interval, see Charter, 1994). A chi-square test, which was used due to the dichotomous nature of this variable, did not indicate a significant difference among groups for the variable BCTChance (X^2 (4, $N = 150$) = 4.07, $p = .402$). Only one participant (NF) actually scored over the cut-off (see Table 19). SPMChance was calculated for SPM as any score less than 13 for the total test. This was a combination of the 95% confidence intervals for both subtests A through B (score < 6) and C through E (score < 7; see Charter, 1994 for a description of confidence interval estimation). A chi-square test demonstrated significant differences among groups for SPMChance (X^2 (4, $N = 150$) = 16.32, $p = .003$; see Table 19). None of the 90 BC, BN, or CN participants scored below chance, whereas 6 of the 60 participants in the faking groups did.

Hypothesis 3. The third hypothesis examined the efficacy of the performance curve strategy for the detection of malingering. Specifically, the fakers (NF, SF) were predicted to evidence a smaller rate of decay (SPMRD) across the SPM subtests than non-fakers as a result of missing a disproportionate number of easy items. Since the assumption of normality was confirmed for SPMRD, parametric statistics were utilized to

test the hypothesis that faking groups would produce a lower SPMRD. ANQ and IQ were significantly correlated with SPMRD (r =.267, r =-.324, respectively, $p < .001$ for each), and were used as covariates in an ANCOVA. The effect of ANQ was not significant for the model (F (1, 143) =.01, p =.922), while the effect of IQ was significant (F (1, 143) = 17.486, $p < .001$). Group means for SPMRD were found to differ significantly (F (6, 143) = 9.73, $p < .001$; see Table 16). The planned comparisons test revealed that the hypothesis was not supported (F (1, 143) = 0.35, p = .545, see Table 20). Further examination of the group means and post-hoc analyses with Tukey HSD revealed that the group differences were not all in the predicted direction (see Table 21). While the NF group had a smaller mean SPMRD than either the BC (p = .001) or the BN (p = .002) groups, the mean SPMRD for SF was not significantly different than that of the BC or BN groups. Unexpectedly, the CN group had the lowest mean SPMRD of all five groups and was significantly different than SF (p = .005). Tukey HSD yielded three homogeneous subsets on SPMRD: a) CN and NF; b) NF and SF; and c) SF, BN, and BC.

Hypothesis 4. The fourth hypothesis examined the efficacy of the magnitude of error strategy to detect malingering on SPM, predicting that fakers would produce a higher number of gross errors than BC, BN and CN participants. Gross errors were empirically derived in this study and defined as those answers which were farthest from correct and were not likely to be chosen by individuals giving their best effort. SPM Gross Errors (SPMGE) were determined by an item analysis of the responses of the BN and CN groups. Responses which were given by less than 5% of the participants were included in SPMGE. Infrequent responses included 44 on subtest A, 47 on subtest B, 64

on subtest C, 54 on subtest D, and 29 on subtest E, for a total of 238 responses on 57 items. Gross errors were calculated for each participant and a cut-off score was determined.

Since the distribution of SPMGE violated the assumption of normality, the Kruskal-Wallis, a nonparametric version of the one-way analysis of variance for independent samples, was utilized to test for differences between group means. Significant differences were found for SPMGE ($\underline{X}^2$ (4, $\underline{N}$ = 150) = 81.41, $\underline{p}$ < .001; see Table 14). Group comparisons were also completed using an ANCOVA with ANQ as a covariate, and significant differences were again identified for SPMGE ($\underline{F}$ (5, 144) = 25.81, $\underline{p}$ <.001; see Table 16). Planned comparisons tests supported the hypothesis that faking groups scored significantly higher on SPMGE than did non-faking groups ($\underline{p}$ < .001 for all comparisons, see Table 22). Since the results of the Kruskal-Wallis, ANCOVA, and planned comparisons tests were similar, further post-hoc analyses with Tamhane's T2 for unequal variances were completed to examine where differences lay. The NF and SF group means were significantly higher than BC, BN, and CN ($\underline{p}$ < .003 for all comparisons, see Table 21), supporting the hypothesis that fakers would score significantly more gross errors than non-fakers. NF and SF did not differ from one another on mean SPMGE ($\underline{p}$ = .153). In addition, the mean for CN was significantly lower than for BC ($\underline{p}$ = .003) and BN ($\underline{p}$ = .001), which did not differ from one another ($\underline{p}$ = 1.00).

Hypothesis 5. The last hypothesis examined the efficacy of the SPM measure of discrepancy (Raven, 1977) to detect possible malingering. The NF and SF groups were

expected to have higher mean numbers of discrepant scores on the SPM than the BN, BC, and CN groups. Individual subtests of the SPM were scored by hand for discrepancies (SPMDIS) according to Raven (1977) instructions. A participant's score for each subtest was subtracted from a score derived from a table which listed the normally expected score for each set according to the total score on the entire scale. A difference of 2 or more between the observed and expected scores on any subtest constituted a discrepancy. Each participant received an SPMDIS score of 0 if there were no discrepancies, a score of 1 if there was one or more discrepancies, and a score of 2 if the total SPM score fell below 15. Group differences on SPMDIS were then examined using a chi-square test of significance for categorical data. A significant difference was detected between groups ($\underline{X}^2$ (8, $\underline{N}$ = 150) = 43.65, $\underline{p} < .001$; see Table 19). In each of the BC and BN groups, 24 (80%) participants had a score of 0, 5 (17%) had a score of 1, and 1 (2%) had a score of 2 (see Table 19). The CN group produced 29 (97%) protocols with an SPMDIS score of 0; only 1 CN participant had a discrepant protocol. Half of the NF group had discrepant protocols, while another 7 (23%) had unscorable protocols. The rest of the NF group (27%) had nondiscrepant protocols. The SF group contained 2 (7%) unscorable protocols, 17 (57%) without discrepancies and 11 (37%) with discrepancies. Therefore, differences were in the predicted direction.

Discriminant Functions Analyses

A direct discriminant function analysis was performed using 5 comparison variables (TBCT, TSPM, BCTTSVI, SPMVI, and SPMGE) as predictors of membership in two groups (faking vs non-faking). Since the BCT VI's were highly intercorrelated,

only one was used in the analysis. BCTTSVI was chosen because it had been proven to be a good predictor in a previous study (Tenhula & Sweet, 1996).

Two separate discriminant function analyses were completed in this study. The first was designed to distinguish between brain-injured and faking participants. However, since the main goal of the study was to distinguish simply between fakers and non-fakers, separate discriminant functions were conducted which compared BC, BN, and CN to the faking groups. When the brain-injured and control groups were combined, there was a slight increase in the overall correct classification rates achieved by the analysis as opposed to when the groups were analyzed without the controls. The increase in classification was likely due to the controls maximizing the difference between the means of the three non-faking groups combined and those of the faking groups combined. Consequently, results from both these analyses are discussed.

The total number of participants included in the first set of discriminant functions analyses was 120: the clinical groups (BC and BN) and the faking groups (NF and SF). The groups were of equal sizes (N = 60) and there were no missing data. Box's M test for the equality of the group covariance matrices was significant, indicating unequal variance-covariance matrices. Although discriminant function analysis is robust to mild violations of the assumption of equal within-group variance-covariance matrices, the extent to which these data violated assumptions warrants that the results be interpreted with caution.

Five discriminant functions were developed based on the variables described above (TBCT, TSPM, BCTTSVI, SPMVI, and SPMGE). These functions achieved hit rates

ranging from 60.0% (TBCT) to 75.0% (SPMGE). The sensitivity rates ranged from a low of 48.3% for SPMVI to 68.3% for SPMGE. The overall discriminant function using all five predictors had a hit rate of 86.7%, sensitivity of 68.3%, and specificity of 81.7% (see Table 23).

The total number of participants included in the second group of analyses was 150. The two groups utilized in this analysis were comprised of non-fakers (BC, BN, and CN, N = 90) and fakers (NF, SF, N = 60). The groups were not of equal sizes but there were no missing data. This data set suffered from the same inability to meet all the assumptions for a discriminant functions analysis as were described above. Five discriminant functions were developed based on the variables described above (TBCT, TSPM, BCTTSVI, SPMVI, and SPMGE). These functions achieved hit rates ranging from 65.3% (TBCT) to 80.0% (SPMVI; see Table 24). The discriminant function developed using TSPM had the highest sensitivity (81.7%) but also one of the lower specificities (75.6%). The discriminant function using all five predictors had an overall hit rate of 81.3%, sensitivity of 73.3%, and a specificity of 86.7%.

Debriefing Questionnaire

The Neuropsychological Simulation Debriefing Questionnaire (NSDQ) contained a total of 13 questions, two of which were short answer while the remainder were multiple choice. Responses to the first two questions demonstrated an adequate understanding of the instructions (93% in NF and 100% in SF). When asked what their level of understanding was regarding the instructions, 53% in the NF group and 87% in the SF group endorsed "very good understanding." Only 10% of the SF group felt they had a

"highly specialized" understanding of brain function, while approximately half of each group felt they had a "good" understanding and only 13% of the NF group felt they had a "poor" understanding. Seventy percent of the NF group and 87% of the SF group rated their efforts to fake as "great," while no members of either group endorsed "gave it no effort." Only one person in the SF group reported not attempting to fake one of the measures. In this case the participant related that he had no experience with the SPM, did not know how to go about responding as if he had brain damage, and decided to respond with a normal pattern.

Participants responded that they attempted to fake the BCT by giving the wrong answers (NF 87%, SF 100%), slowing their performance (NF 53%, SF 37%), appearing confused (NF 40%, SF 33%), simulating memory impairment (NF 37%, SF 50%), and ignoring examiners instructions (NF 13%, SF 10%). Only one participant in the SF group attempted to change emotional states as a way of simulating brain injury.

Participants responded that they attempted to fake the SPM by giving the wrong answers (90% for both NF and SF), slowing their performance (NF 40%, SF 17%), appearing confused (NF 37%, SF 23%), simulating memory impairment (NF 13%, SF 17%), and ignoring examiner's instructions (NF 7%, SF 3%). Only one member of the NF group attempted to change emotional states as a way of simulating brain injury. Utilizing Chi-square tests for significance, responses to two questions were different between the groups, with the SF group responding more to "gave the wrong answers" than the NF group ($\underline{X}^2$ (1, $\underline{N}$ = 60) = 4.29, $\underline{p}$ < .038), and the NF group responding more to "slowed my performance" ($\underline{X}^2$ (1, $\underline{N}$ = 60) = 4.02, $\underline{p}$ < .045).

Only 7% of the NF group and 3% of the SF group felt they were "very successful" at convincing the examiner that they were giving a good effort on the measures. Thirty-three percent of the NF group and 57% of the SF group felt they were partially successful in convincing the examiner. When asked how successful their overall attempts to fake were, 43% of the NF group and 77% of the SF group felt they were "partially successful" while 27% of the NF and 17% of the SF group felt they were "partially unsuccessful." Twenty-seven percent of the NF group felt they were "unsuccessful in their attempts to fake," while only one participant in the SF group responded in this manner. Seventy-three percent of the NF and 27% of the SF group did not respond to the question "If you feel you were successful, then what helped you to fake?" Of those who did respond to this question, 50% of the NF group and 81% of the SF group felt that "knowledge of the brain" had assisted them, while 50% of the NF group and only 19% of the SF group felt that having known someone with brain damage contributed to their success. Fifty-three percent of the NF group and 47% of the SF group did not respond to the question, "If you did not feel you were able to fake well, what hindered you?" Of those who responded to this question, 43% of the NF group and 13% of the SF group felt they were hindered by their honesty, 43% of the NF respondents and 25% of the SF respondents felt the test was "too easy," while 13% of the NF respondents and 63% of the SF respondents felt the measures were "too hard."

Exploratory Analyses

Because two clinical brain injury groups and two faking groups were used in this study, exploratory analyses were conducted to compare the groups within each set. BC

and BN differed only with respect to the demographic variables of diagnosis, mean Rey score and mean months post-injury, as reported above. No differences were found between the two clinical groups on any of the comparison variables (see Table 25). Exploratory analyses conducted between the NF and SF groups showed differences on demographic variables for IQ (t (58) = 6.50, $p < .001$), with SF having a higher mean IQ, and for ANQ (t (39.3)[2] = 3.17, $p = .003$), with NF having almost twice as high an ANQ score. Differences were also noted with respect to three comparison variables (see Table 26). NF scored significantly higher than SF on TBCT (t (58) = 2.43, $p = .018$), BCTBVI (t (41.9) = 2.24, $p = .030$), BCTCVI (t (40.9) = 2.22, $p = .032$), SPMVI (t (55.5) = 4.07, $p < .001$), and SPMGE (t (58) = 2.47, $p = .016$). SF scored significantly higher than NF on TSPM (t (58) = -3.03, $p = .004$) and SPMRD (t (58) = -2.05, $p = .045$). NF also had more unscorable SPM protocols (SPMDIS) than did SF (X^2 (4, N = 150) = 6.63, p =.036).

Cut-off scores were empirically derived for each of the variables in this study to provide clinical decision rules and to determine hit rate, sensitivity, and specificity. First, frequency distributions based on the non-faking groups (CN and BN) were examined for each variable. Next, the cut-off score was derived by choosing the number over which less than 5% of these groups scored. Then the cut-off score was applied to each group to determine an overall hit rate (percent of participants who were correctly categorized). Sensitivity rate is defined in this study as the percent of NF and SF who were correctly

---

[2]Degrees of freedom which contain decimal points are for t-tests conducted with samples having unequal variances.

categorized as faking. Specificity rate is the percent of BC, BN and CN group participants who were correctly categorized as non-fakers. (See Table 27 for a listing of hit rates, sensitivities, and specificities of each variable).

The VI's derived and examined in this study yielded less than impressive results. BCTTSVI established the best overall hit rate of 71% with a sensitivity rate of 32% and specificity of 97%. The overall hit rate for BCTBVI was 67% with a sensitivity of 27% and a specificity of 98%. The BCTIVI derived in this study had an overall hit rate of 66% with a sensitivity of 27% and a specificity of 96%. The SPMVI derived in this study had better rates than any of the BCT VI's, with a hit rate of 76%, sensitivity of 48% and a specificity of 94%. In summary, the VI's had higher than chance hit rates and very rarely misclassified non-malingers, but they were lacking in their ability to correctly categorize participants asked to fake.

Clinical decision hit rates for both BCT below chance responding (BCTChance) and SPM below chance responding (SPMChance) were very poor. Few of the faking members actually scored below level of chance on the measures utilized in this study (NF = 6, SF = 1). SPMChance was the better of the two variables at detecting faking with an overall hit rate of 64%, sensitivity of 10%, and specificity of 100%. BCTChance had an overall hit rate of 41%, sensitivity of 2%, and specificity of 100%.

The interpretation of SPMRD was clouded by the curvilinear relationship between RD and Total SPM score, with RD tending to be highest for scores between 30 and 40. In order to increase its interpretability, cut-off scores based on total scores were derived. Each score was assigned to one of ten categories, according to the method used by

Gudjonsson and Shackleton (1986; see Table 28). Then the average rate of decay for the BN and CN groups was calculated in each total score category. No scores were found to fall within the first category, whose total scores ranged from 0 to 12. Next, these average scores were divided in half to determine the SPMRD cut-off scores for each total score category (see Table 28). The cut-off scores were then applied to all groups. Utilizing this method, SPMRD had an overall hit rate of 66% with a sensitivity of 32 % and specificity of 89%.

SPM gross errors (SPMGE) had the highest hit rate of all the variables examined in this study (78%), with a sensitivity of 57% and a specificity of 92%. The SPM discrepancy score had the next best hit rate of 77 % with a sensitivity of 43% and a specificity of 86%.

The total scores for BCT (TBCT) and SPM (TSPM) were also examined for their contribution to the classification of fakers. TBCT achieved an overall hit rate of 65% with a sensitivity of 22% and a specificity of 93% while TSPM had a hit rate of 72%, sensitivity of 43% and a specificity of 86%.

The SF group evaded detection at a higher rate than the NF group across the board (see Table 29). The highest rates of detection for the SF group were on SPM (SPMGE and SPMDIS), indicating that the SPM may be more difficult to fake than the BCT. The rates of detection for TSPM were somewhat lower (specificity NF = 40%, SF = 23%) than for SPMVI (specificity NF = 67%, SF = 30%).

Exploratory analyses were conducted utilizing discriminant function analyses for the individual measures BCT and SPM between the groups non-fakers (BC, BN, and CN)

and fakers (NF, SF). A discriminant function utilizing predictors specific to the BCT (TBCT, BCTTSVI) had a hit rate of 71.3%, sensitivity of 63.3% and specificity of 76.7%. A discriminant function utilizing predictors specific to the SPM (TSPM, SPMVI, and SPMGE) attained a hit rate of 80.7%, sensitivity of 76.6% and a specificity of 83.3% (see Table 30).

A separate set of discriminant functions was completed for the BCT and SPM between the clinical (BC, BN) and faking groups (NF, SF). A discriminant function utilizing predictors specific to the BCT (TBCT, BCTTSVI) had a hit rate of 68.3%, sensitivity of 60%, and specificity of 76.7%. A discriminant function utilizing predictors specific to the SPM (TSPM, SPMVI, and SPMGE) attained a hit rate of 76.7%, sensitivity of 66.7%, and specificity of 86.7%. Overall, the SPM seemed to have the highest rates of detection across groups.

CHAPTER IV

DISCUSSION

In general the hypotheses that fakers would perform worse on two neuropsychological measures were supported. When a validity index comprised of infrequently missed items was applied to fakers, their scores tended to be higher than those of the non-fakers. Likewise, the faking groups gave more gross errors and produced more discrepant protocols than the non-faking groups. Chance responding and rate of decay across SPM subsets were not found to be useful in distinguishing between fakers and non-fakers.

The hypothesis that fakers would fail more obvious items than the other groups (BC, BN, and CN) was not supported although fakers scored higher than all other groups except BN on most VI's. Specifically, the SF participants were consistently able to produce protocols which were not significantly different from those of the BN. Of the 5 validity indexes used to examine the BCT, the BCTTSVI (Tenhula & Sweet, 1996) and BCTIVI (derived from this study) were the most efficient at detecting faking, with sensitivity rates of 32% and 27%, respectively. When the SF participants were removed from the analysis, sensitivity did not increase dramatically. Tenhula and Sweet (1996) reported much higher rates of detection in their study (76%). One reason for the difference may be that their group of fakers was comprised of undergraduate college students who were completing the BCT for extra credit while this study utilized a

community sample of simulators. Another reason for the difference may be that Tenhula and Sweet (1996) include instructions to fake which may have lead their faking group to overextend their performance (i.e., "Simulators were encouraged to fake the most severe disability that they could without letting the examiner know that they were faking" p. 108), while this study encouraged faking "as if you had a brain-injury" without any reference to severity. Finally, differences between the detection rates of BCTTSVI and BCTIVI may be attributable to the number of items on each index; the BCTTSVI is comprised of 19 items while the BCTIVI is comprised of only 13. This is confirmed by the general trend among the VI's for those with fewer items to have lower sensitivity rates (e.g., the rate for BCTTSVI with 19 items was 32%, while the rate for BCTCVI with 6 items was 19%). Overall, the sensitivity rates for the VI's were extremely low, indicating their vulnerability to false positives.

Contrary to the current study, which utilized Bolter's more recent VI (1992), and Tenhula and Sweet (1996), which used Bolter's initial VI (1985), Trueblood and Schmidt (1993) did not find that suspected malingerers made more frequent errors than controls on Bolter's VI (1985). A different cut-off score might have increased the detection rate in Trueblood and Schmidt's (1993) study (Tenhula & Sweet, 1996). In light of the low sensitivity rate using Bolter's items in this study and the lack of significant results by Trueblood and Schmidt (1993) with an identified malingering group, the clinician should use caution when drawing conclusions based on the this VI alone.

The SPM Validity Index (SPMVI), had an overall hit rate which was somewhat higher (76%) than the VI's for BCT. This was especially true for detecting NF

participants who were lower than all other groups, while SPMVI was not as good at detecting SF participants. One explanation may be that SPM lacks performance feedback, which is integral to the BCT. With a lack of feedback, participants are not as aware of how many items they have missed, therefore they may miss more items overall. Another explanation may be the inability of participants to reliably predict the increasing difficulty of items both within and across the subtests. As a result, fakers may miss more of the easy items without knowing how difficult the later items will become. No validity index has previously been developed for the SPM, and the SPMVI offers the clinician a useful tool in discerning test-taking effort.

Overall, the below chance variables were very poor measures for the detection of faking in this study. The hypothesis that fakers would score below chance more frequently than non-fakers was not significant for BCT but was for SPM, and the specificity rate for both was 100%. However, these are misleading results because only 7 of the 60 participants in the faking groups were actually detected using this method, and the better of the two measures, SPMChance, had a very low sensitivity rate of only 10%. This is a case in which statistical significance does not mean clinical significance. Perhaps these two measures do not lend themselves well to an evaluation of chance responding.

The results of this study did not completely support the hypothesis that rate of decay would be reduced for fakers. In fact, the group with the lowest rate of decay was the CN group. The overall hit rate for the SPMRD in this study was 66%. In contrast, Gudjonsson and Shackleton (1986) reported a hit rate of 83%. The differences between hit rates in these two studies may be a result of differences in IQ. The brain-injured

participants in the Gudjonsson and Shackleton (1986) study had a lower mean IQ than their counterparts in the present study. RD in individual cases should be interpreted with reference to total score and overall IQ, and RD is not as useful with people who tend to score within the very low or very high range on SPM (e.g., total scores of 0 and 60 would both have a 0 rate of decay). The hit rates may also have been affected by the difference between the faking instructions. Gudjonsson and Shackleton's (1986) participants were asked to "fake substantially and convincingly below their genuine ability on the tests..." (p. 36), with no mention of possible repercussion for being detected as malingering. In contrast, faking participants in the present study were warned of the result of being detected. Lastly, there is certainly a difference in group composition between the two studies. While the non-clinical groups in this study were comprised primarily of community samples (excluding SF), the Gudjonsson and Shackleton (1986) groups contained British soldiers, including nurses and bomb disposal experts.

The hypothesis that fakers would make more gross errors on SPM than non-fakers was fully supported. SPMGE allowed for the highest hit rate and sensitivity of all variables examined in this study (78% and 57%, respectively). Yet SPMGE, with 238 responses qualifying as gross errors, is limited in its practicality by the amount of time required for hand scoring. A computer program which incorporates the scoring of gross errors as responses are entered would increase the utility of SPMGE. Exploratory analysis indicated that the inclusion of SPMVI as a second measure of possible faking enhanced the probability of detecting faking.

The hypothesis that fakers would make more discrepancy errors (SPMDIS) on the SPM was also fully supported in this study. The overall hit rate for SPMDIS was 77%, which was very close to the 74% found by Gudjonsson and Schackleton (1986). Although the hit rate was moderately high, SPMDIS was more prone to false positives than the other variables examined in this study (15%). TSPM had very nearly the same rates of detection as did SPMDIS, indicating that the clinician may not need to determine whether a discrepancy has occurred but could use the recommended cut-off for the total score and achieve similar rates of accuracy.

Although no specific hypotheses were made concerning differences between BC and BN groups or NF and SF groups, BC was expected to score worse than BN due to increased motivation (i.e., involvement in litigation and chance for compensation), and NF was expected to score worse than SF due to their lack of sophistication regarding training in and knowledge of brain functioning. However, the anticipated differences between the BC and BN groups did not materialize; no significant differences were found on any of the comparison variables, but the BN group did show a trend to perform somewhat worse than BC, with most means, although not significant, evidencing poorer performance. The BN group may have looked more impaired for several reasons. First, they scored higher on the ANQ, evidencing more reported neurological symptomology. Next, the significant difference between the groups for diagnosis evidenced more closed head injuries (CHI) in the BN group and more post-concussive syndrome (PCS) in the BC group. The participants with CHI may have had more devastating and lingering injuries than the participants with PCS. More months post-injury indicates more chronic or longer-term

and lingering deficits in the BN group. The significantly lower score on the Rey by the BN group is also evidence that these participants likely had more impairment in memory functioning than the BC participants.

The differences between the NF and SF groups were consistent with expectations that SF participants would score more similarly to the clinical groups and be better at avoiding detection than NF. Significant differences on the comparison variables were found for TBCT, TSPM, BCTBVI, BCTCVI, SPMVI, SPMRD, and SPMGE, with NF scoring worse than SF. In addition, specificity rates for SF were lower across the board, indicating better ability to avoid detection. Level of prior exposure to these tests was not examined in this study but the SF participants likely had more experience with the BCT than the SPM. Since neither NF or SF participants were coached on specific symptoms before completing the measures, the lower specificity rates for SF provide evidence that exposure to neuropsychological tests and information on basic brain functioning can contribute to lower detection rates.

According to the debriefing questionnaire, most of the malingerers attempted to fake by employing more than one strategy. Among these, the most frequent approaches were giving the wrong answers, slowing performance, appearing confused, simulating memory impairment, and ignoring examiner's instructions. Only one person (in the SF group) attempted to change emotional states as a way of simulating brain injury. The only differences between the NF and SF groups in their faking strategies were the SF group responding more to "gave the wrong answers" and the NF groups responding more to "slowed my performance." More of the SF group (77%) felt they were "partially

successful" in faking than the NF group (43%), while only one person in both the SF and NF felt they were "very successful" in faking.

Strengths of the Study

Well-designed research has the goal of empirically answering theoretical questions while maximizing the effects of independent variables, minimizing the effects of random error, and controlling the effects of extraneous variables. In the present study, the primary independent variable of interest was group membership, and several techniques were used to create unique groups. For example, in order to address possible motivational issues regarding litigation and compensation, two different clinical groups were included, one which was involved in litigation and one which was not. The SF group was incorporated to examine the prediction that being familiar with tests of brain functioning and basic neuropsychology would influence detection rates. Rogers' (1988) recommendations were also incorporated to help maximize the effects of group: a) incentives were provided which included financial reward for eluding detection, b) faking instructions were precise and believable without encouraging the fakers to overdo their performance; c) a debriefing questionnaire was used to assess the accuracy of compliance with instructions, and d) multivariate analyses were utilized when appropriate.

Random error was minimized in several ways. First, tests administered were commonly used and reliable neuropsychological measures. Second, although the testing settings varied by groups, the experimental setting was as free as possible from distractions. In addition, the battery was short enough that fatigue was not an issue, with the exception of the BC participants, who completed a full neuropsychological battery.

Researchers strive not only to maximize differences between groups with respect to dependent variables, but also to control for effects of extraneous variables, such as age, gender, intellectual functioning, etc. In the present study, membership in three groups (BC, BN, and SF) was predetermined by participant history. However, random assignment was used with the remaining two groups (CN and NF) in order to equalize the effects of extraneous variables. In addition, a community sample, rather than college students (e.g., Tenhula & Sweet, 1996), was used for the CN and NF groups, thereby diversifying the sample, more closely matching groups with respect to age, SES, intellectual functioning, and education level, and increasing external validity. The use of relatively homogeneous college student samples has been criticized for its lack of similarities to actual brain-injury litigants (Nies & Sweet, 1994).

In order to reduce the threats to internal validity, the present study was designed to maximize power where possible. With the multiple comparison variables and five different groups, the number of possible comparisons was unwieldy, and the chance for spurious findings was great. Theory-based research reduces the number of comparisons by making predictions about expected differences and analyzing only those variables. Thus, the number of comparisons was minimized through the use of a priori hypotheses and planned comparisons when possible. Power was also increased through the use of validity and gross error scales which were derived based on both clinical (BN) and control (CN) groups; thus the VI's were a truer reflection of the performance of participants with brain injury and without.

Another strength of the study was the cautious selection of statistical measures based on the nature of the underlying distributions of variables. Parametric statistics such as ANOVA and MANOVA were only applied if underlying distributions were normal or if results with the non-parametric measures were highly significant, in which cases the parametric statistics showed similar highly significant between-group differences.

Limitations of the Study

Group differences with regard to demographic variables may have contributed to some of the unexpected findings. The use of a higher number of CHI participants in the BN group, who also had a greater number of months post-injury and more neurological damage than the BC group, may have contributed to the poorer performance of BN compared to BC. In addition, the BC group was being evaluated in an office setting as part of a lengthy battery of measures which were planned to be used in litigation for compensation, while the BN participants were tested under different conditions, solicited as volunteers in the experiment with no compensation. As a result, the BN participants may have been less than optimally motivated to perform well. Uniform collection of data could reduce these mitigating effects of random error in future research.

A significant weakness of most simulation studies is the inability to offer incentives commensurate with personal injury awards (Rogers, 1997). The low-level inducements utilized in this study ($15.00 for participation and a chance for $50.00 if detection was avoided) are not directly comparable to the potential monetary value which may be awarded in real-world cases, but are representative of the financial constraints involved in

conducting research. However, some studies have found no difference between simulators based on level of inducement to malinger (e.g., Bernard, 1990; Binder & Willis, 1991).

Clinical evaluations of brain-injured patients usually involve a battery of tests, but, due to the demands on time and effort, the full-battery approach is not practical with volunteer participants. Yet patterns of responding on an entire battery may differ from those on one or two tests, and faking brain impairment on a single test could be easier than faking throughout an entire battery of tests (Nies & Sweet, 1994). Therefore, this study, while using more than one measure, falls short of a typical neuropsychological test battery, so the degree to which these findings generalize is unknown.

Random assignment to CN and NF groups was used to control for extraneous variables and to allow researchers to be blind to the group membership so as not to influence responding. However, group membership was sometimes apparent as evidenced by test-taking attitude, number of items missed, and exhibited effort. For example, most control participants displayed signs of disappointment or frustration when items were missed, but faking participants in general did not. Researchers made every effort to administer tests in a uniform matter regardless of the test-taking attitude of the participants.

Perhaps the major limitation of the present study was the within-groups distributions of the comparison variables, which were typically asymmetric with heterogeneous variances. The distributions were positively skewed on some variables for some groups and negatively skewed on the same variables for other groups. As a result, data transformations did not symmetrize the distributions among variables. Although

nonparametric and parametric statistics yielded similar strong between-groups differences, the violations of numerous assumptions associated with parametric statistics call into question their appropriateness with this data. However, each comparison variable utilized in this study was chosen because of an a priori hypothesis regarding its ability to distinguish between groups of fakers and non-fakers. Cut-offs were generated for these variables which automatically placed the distributions for groups at opposite ends. As a result, any research using these and similar techniques will be plagued with statistical complications. At the very least, results of parametric tests with non-normally distributed data should be interpreted with caution.

Suggestions for Future Research

Future research on malingering would benefit from several modifications of the present study. A group of actual malingerers could be identified with a preliminary measure which incorporates a technique such as Symptom Validity Testing. In the present study, malingering status could not be guaranteed in the groups which were asked to fake even though a self-report was utilized. Uniform testing environment across groups would help to minimize random error. Group differences between litigants and nonlitigants might be maximized if groups were matched according to diagnosis and months post-injury. Increasing incentives for fakers, as well as the chance for a loss of incentives due to detection, may make their performance more similar to actual malingerers, thereby extending the generalizability of simulation results. For example, instead of the common practice of offering an extra monetary incentive when detection is avoided, researchers

may make the basic monetary incentive strictly contingent upon the avoidance of detection.

In future studies, the application of the most promising methods of detection from this study (SPMVI and SPMGE) to protocols of identified malingerers could yield important information about the generalizability and clinical usefulness of these measures. The method of gross error estimation utilized in this study could also be expanded for use with other neuropsychological measures.

The examination of different levels of sophistication of knowledge of brain functioning and injury is another area for future research. Although the SF participants in this study may be considered to be highly sophisticated compared to most people, some malingerers who have researched information about brain functioning or have been coached on how to respond may possess a similar level of sophistication as the graduate students. Researchers may also study how sophistication is inadvertently gained through repeated medical, psychological and neuropsychological interviewers' questioning of the sequelae of brain injury. Perhaps the frequent interviews contribute to an individual's ability to know which symptoms he or she should elaborate or fake.

The examination of physiological responses during testing may also hold promise for future research in the detection of malingering (Nies & Sweet, 1994). For example, recordings of the P300 event-related brain potential response have yielded strong correlations between behavioral performance and brain functioning (Rudell, 1991; Verlager & Berg, 1991; and Rosenfeld, Sweet, Chuang, Ellwanger, & Song, 1996). The

ability to find discrepancies between a verbal response and a brain response is compelling and deserves further exploration.

Another possible area of future research is the development or extension of structured interviews to the detection of malingering of neuropsychological deficits. Although not yet utilized with a neuropsychological population, structured interviews (e.g., Rogers, Gillis, Bagby, & Monteiro, 1991; Rogers, Gillis, Dickens, & Bagby, 1991) have been useful in the detection of malingering with other populations.

In summary, systematic ways of analyzing and attempting to detect malingering by psychological methods began after World War II (Nies & Sweet, 1994). In the more than 50 years since this time, neuropsychology has established multiple new instruments including computer analysis and physiological monitoring. The majority of the measures being used today in the detection of malingering have been created within the past 5-10 years (Nies & Sweet, 1994). Although neuropsychology has come a long way in developing measures to detect malingering, improvements still need to be made.

APPENDIX A

NEUROPSYCHOLOGICAL SIMULATION

DEBRIEFING QUESTIONNAIRE

Appendix A

Neuropsychological Simulation Debriefing Questionnaire

What do you think the purpose of this experiment was?

_____

_____

_____

Would you please paraphrase the instructions you received at the beginning of this assessment?_____

_____

_____

What was your level of understanding regarding the instructions?

a. very good understanding
b. limited understood
c. moderate understanding
d. did not understand

At what level would you rate your understanding of the instructions to fake?

a. very good understanding
b. understood most
c. understood some
d. very little understanding

At what level do you think your knowledge of brain function is?

a. highly specialized
b. good
c. adequate
d. poor
e. _____

How would you rate your efforts to fake these measures?

a. great effort
b. gave it a half effort
c. little effort
d. gave it no effort
e. _____

Which tests did you attempt to fake?

a. 1st
b. 2nd
c. Both

In faking the 1st test, what did you do in trying to fake? (Check all that apply)

_____gave the wrong answers
_____slowed my performance
_____tried to appear confused or unable to comprehend
_____simulated memory impairment
_____stuttering
_____ignored examiners instructions
_____changed emotional state
Elaborate:

_____
_____
_____

In faking the 2nd test what did you do in trying to fake?
(Check all that apply)

_____gave the wrong answers
_____slowed my performance
_____tried to appear confused or unable to comprehend
_____simulated memory impairment
_____stuttering
_____ignored examiners instructions
_____changed emotional state
Elaborate:

_____
_____
_____

How successful do you believe you were at fooling the examiner?

a. very successful
b. partially successful
c. partially unsuccessful
d. unsuccessful
e. _____

How successful do you believe your attempts to fake were?

a. very successful
b. partially successful
c. partially unsuccessful
d. unsuccessful
e. _____

If you feel you were successful, then what helped you to fake?

a. Knowledge of brain
b. Known people with brain damage
c. _____

If you did not feel you were able to fake well, what hindered you?

a. I am too honest
b. I didn't understand the directions
c. Too easy
d. Too hard
e. _____

APPENDIX B

INFORMED CONSENT

Appendix B

Informed Consent Form

Investigators: Primary Researcher: Chuck Isler, M.A.
Faculty Researcher: E. H. Harrell, Ph.D.

Purpose of the Study: The purpose of this study is to examine the performance of brain-damaged (litigation v non-litigation) versus non-brain-damaged individuals on two neuropsychological tests and how faking the tests might differentiate the groups.

Procedures: The participant will attend one session of neuropsychological testing lasting approximately two hours.

Safeguards: The participant's answers will be kept confidential. All participant information and scores will immediately be assigned a code number. Individual responses will be pooled into group means for analyses. A master list will be kept with the name of the participant and research number for identification of those who are able to elude detection.

Participation in this study is entirely voluntary. You may end your participation at any time you desire. To the best of our knowledge, participation will not cause any physical or psychological harm.

Benefits: This study provides an opportunity for participants to increase their knowledge and understanding of the performance of brain-damaged individuals. This study may potentially benefit society in general and handicapped individuals in particular as we learn more about the challenges and possible ways to treat their problems.

Additional Questions: If you have questions regarding your participation, or this study, please direct them to Chuck Isler or E. H. Harrell at (817) 565-2671.

Informed Consent Form

This is to certify that I _____
(your name)
am participating as a volunteer in a scientific study as an authorized part of the educational and research program of the University of North Texas.

This study as well as my participation in the investigation have been defined and fully explained to me by the researcher. The procedures of this study and their possibly risks and discomforts have been described in a separate statement, and are discussed in detail.

I have been given an opportunity to ask any questions I may have, and all such questions and inquiries have been answered to my satisfaction;

I understand that I may refuse to answer any questions on the questionnaires.

I understand that any data or answers to questions will remain confidential and I will not be able to view my individual results. Only group data will be reported.

I also understand that I am free to withdraw my consent and end my participation at any time during the study without penalty, prejudice, or loss of benefits.

_____
(date)

_____
(signature)

This project has been reviewed and approved by the UNT Committee for the Protection of Human (817) 565-3940.

APPENDIX C

TABLES

Table 1

Comparison of Participants' Gender, Handedness, Ethnicity, and Income by Group.

Gender	BC	BN	CN	NF	SF	Total	X^2	d.f.	p
Male	16	22	9	9	11	67	16.89	4	.002
Female	14	8	21	21	19	83			
Handedness									
Right	25	29	26	29	27	136	5.04	4	.283
Left	5	1	4	1	3	14			
Ethnicity									
Caucasian	26	27	29	27	26	135	12.01	16	.743
Hispanic	2	2	0	1	2	7			
Afr-Amer	1	0	1	0	1	3			
Asian	1	1	0	2	0	4			
Other	0	0	0	0	1	1			
Income									
$0-10K	9	9	5	8	13	44	23.61	24	.484
$10-20K	3	9	5	6	7	30			
$20-30K	7	4	4	6	1	22			
$30-40K	6	3	8	7	5	29			
$40-50K	1	1	3	1	2	8			
$50K+	4	4	5	2	2	17			

Note. N=30 for all groups.

Table 2

Comparison of Age of Participants by Group.

Group	Mean	S.D.	Min	Max
BC[a]	42.53	13.46	21	71
BN[b]	39.63	12.30	19	69
CN	37.13	12.00	20	60
NF	32.83	13.47	18	62
SF	29.33	4.27	23	40
Total	36.29	12.41	18	71

$F_{(df=4)}=6.169$, $p<.001$

[a]BC mean age is significantly greater than mean ages for NF ($p=.038$) and SF ($p=.001$).

[b]BN mean age is significantly higher than mean age for SF ($p=.022$).

Table 3

Comparison of ANQ Scores by Group.

Group	Mean ANQ	S.D.	Min ANQ	Max ANQ
BC[a]	20.70	9.54	3	33
BN[a]	16.73	6.13	3	31
CN[b]	6.07	3.90	0	16
NF[b]	6.10	4.76	0	22
SF[b]	3.10	2.04	0	7
Total	10.54	8.99	0	33

Note. $F=52.021$, $p<.001$

[a]ANQ scores for BC and BN were similar to each other and significantly higher than ANQ scores for CN, NF and SF ($p<.001$ for all comparisons).

[b]ANQ scores for CN, NF, and SF were similar to each other.

Table 4

Comparison of Mean IQ Scores by Group.

Group	Mean IQ	S.D.	Min IQ	Max IQ
BC[a]	91.73	10.75	75	118
BN[b]	94.37	17.06	65	133
CN[c]	107.13	13.96	75	131
NF[d]	102.63	11.42	81	125
SF[e]	122.17	11.86	104	145
Total	103.61	16.96	65	145

Note. $F(4, 145) = 25.103$, $p < .001$; Tukey HSD was used for all post-hoc analyses.

[a]IQ scores for BC were similar to those for BN, but significantly lower than those for CN, NF, and SF.

[b]IQ scores for BN were not significantly different than those for BC or NF, but were lower than those for CN and SF.

[c]IQ scores for CN were not significantly different than NF but were higher than BC, BN and lower than SF.

[d]IQ scores for NF were not significantly different than those for BN or CN but were higher for BC and lower for SF.

[e]IQ scores for SF were significantly higher than all other groups.

Table 5

Diagnoses for Clinical Groups.

	Group					
Diagnosis	BC	BN	Total	X^2	d.f.	p
Closed Head Injury	15	25	40	11.11	5	.049
Post-concussive Syndrome	8	3	11			
Cerebral Vascular Accident	2	0	2			
Anoxia	1	2	3			
Toxic Exposure	3	0	3			
Electrical Shock	1	0	1			
Total	30	30	60			

Table 6

T-test Comparisons Between Clinical Brain Injury Groups for Mean Rey Score and Months Since Injury.

	BC		BN		t-test		
	Mean	SD	Mean	s.d.	t	df	p
Rey	13.03	2.36	11.6	2.59	2.24	58	.029
Months since injury	34.3	37.22	93.97	83.76	3.57	40.02	.001

Note. Equal variance was assumed for Rey but not for Months since injury.

Table 7

Group Comparisons with ANOVA for Mean TBCT and TSPM.

	Group Means						ANOVA	
	BC[a]	BN[a]	CN[b]	NF[c]	SF[d]	Total	F	p
TBCT	69.27	67.07	42.20	94.30	75.63	69.69	12.012	.000
TSPM	40.80	38.50	51.83	22.57	30.53	36.85	31.442	.000

Note. Poorer performance is associated with higher scores on TBCT and lower scores on TSPM.

[a]TBCT scores for BC and BN were similar to each other and significantly higher than TBCT scores for NF but lower than CN. TSPM scores for BC were similar for BC and BN, higher than NF and SF and lower than CN.

[b]TBCT and TSPM scores for CN were significantly higher than all other groups.

[c]TBCT scores for NF were similar to SF but lower than all other groups. TSPM scores for NF were significantly lower than all other groups.

[d]TBCT scores for SF were similar to BC, BN, and SF but significantly lower than SF. TSPM scores for SF were similar to BN but significantly lower than BC and CN and higher for NF.

Table 8

Post-Hoc Analyses of Group Differences on TBCT and TSPM.

	BC		BN		CN		SF	
	I-J	p	I-J	p	I-J	p	I-J	p
TBCT[a]								
NF	25.03	.009*	27.23	.003*	52.10	.000*	18.67	.104
SF	6.37	.920	8.57	.796	33.43	.000*		
CN	-27.07	.004*	-24.87	.010*				
BN	-2.20	.999						
TSPM[b]								
NF	-18.23	.000*	-15.93	.000*	-29.27	.000*	-7.97	.045
SF	-10.27	.013*	-7.97	.139	-21.30	.000*		
CN	11.03	.001*	13.33	.000*				
BN	-2.30	.999						

Note. I-J refers to the mean difference between scores for the groups listed in the column on the left (I) and the groups listed across the top (J). For example the mean TBCT score for NF (I) is 94.30 and BC (J) is 69.27 while the difference (I-J) is 25.03.

[a]Homogeneity of variance assumption was met for TBCT. Tukey HSD was used for comparisons.

[b]Homogeneity of variance assumption was not met for TSPM. Tamhane T2 was used for comparisons.

*Signifies those p values which are significant less than .05.

Table 9

Tests for Normality of Distribution of Comparison Variables

	Group									
	BC		BN		CN		NF		SF	
Variable	S-W[a]	p[b]	S-W	p	S-W	p	S-W	p	S-W	p
TBCT	.973	.67	.971	.60	.919	.04	.983	.90	.975	.69
TSPM	.853	.01	.944	.18	.927	.05	.919	.03	.941	.14
BCTBVI	.453	.01	.547	.01	.404	.01	.834	.01	.802	.01
BCTTSVI	.802	.01	.740	.01	.597	.01	.912	.02	.906	.01
BCTIVI	.588	.01	.539	.01	.187	.01	.863	.01	.839	.01
BCTCVI	.277	.01	.430	.01	n/a	n/a	.749	.01	.638	.01
BCTCPVI	.484	.01	.563	.01	.404	.01	.807	.01	.855	.01
SPMVI	.507	.01	.699	.01	.187	.01	.920	.04	.764	.01
SPMRD	.973	.66	.975	.71	.935	.08	.983	.89	.971	.61
SPMGE	.777	.01	.860	.01	.632	.01	.940	.12	.912	.02

[a]S-W refers to the Shapiro-Wilk statistic which is appropriate for sample sizes less than 50. N for each group is equal to 30.

[b]Reported p-values of .01 represent the upper bound of the true significance.

Table 10

Tests for Homogeneity of Variance of Comparison Variables.

	Levene's	
Variable	F	p
TBCT	0.57	.687
TSPM	7.50	.000
BCTBVI	16.45	.000
BCTTSVI	9.84	.000
BCTIVI	11.81	.000
BCTCVI	17.45	.000
BCTCPVI	13.01	.000
SPMVI	29.22	.000
SPMRD	0.32	.866
SPMGE	11.62	.000

Note. d.f.1 = 4 and d.f.2 = 145 for all variables. A significant result means variances are not equal between the groups.

Table 11

Correlations Between Demographic and Comparison Variables.

Comparison	Demographic Variable			
Variable	Age	ANQ	IQ	Gender[a]
TBCT	.149	.020	-.202 (.013)	.001
TSPM	.044	.130	.128	.009
BCTBVI	-.055	-.153	-.034	.120
BCTTSVI	-.020	-.165	-.004	.105
BCTIVI	-.100	-.169	-.024	.132
BCTCVI	-.042	-.159	-.002	.098
BCTCPVI	-.076	-.150	-.003	.133
SPMVI	-.010	-.203	-.023	.095
BCTChance[a]	.124	.069	.055	.114[b]
SPMChance[a]	.133	.103	.003	.073[b]
SPMRD	-.031	.267 (.001)	-.324 (.000)	.116
SPMGE	-.062	-.222 (.006)	-.001	.045
SPMDIS[a]	.143	.163	.093	.099[b]

Note. Significant p values are in parentheses. A cut-off of $p < .01$ was used.

[a]Eta was used as a measure of association except where noted. No p-values are given with eta.

[b]Value reflects the contingency coefficient for nominal by nominal data.

Table 12

Percentage of Correct Answers on Booklet Category Test for BN and CN Groups.

	BCT Subtest						
Item #	I	II	III	IV	V	VI	VII
1	100	92	27	62	25	60	80
2	100	98	43	62	40	80	93
3	100	100	45	70	37	72	90
4	96	100	40	82	38	100	77
5	100	100	30	85	57	65	38
6	96	95	40	83	67	80	95
7	100	100	48	78	38	70	85
8	100	100	51	65	70	72	75
9		100	68	38	62	63	87
10		100	43	75	75	72	95
11		98	77	53	78	75	83
12		100	72	78	78	73	65
13		98	45	67	73	68	97
14		98	63	87	82	78	58
15		95	50	58	82	92	33
16		98	52	77	82	92	95
17		97	60	78	68	83	88
18		98	48	68	28	88	47
19		98	68	85	50	88	82
20		100	58	70	70	95	90

Table Continues

Table 12 - Continued

Item #	I	II	III	IV	V	VI	VII
21			33	82	77	95	
22			67	83	67	92	
23			57	80	80	86	
24			55	72	80	93	
25			60	80	87	95	
26			68	75	80	97	
27			62	70	90	95	
28			67	75	93	93	
29			57	78	87	90	
30			58	65	95	87	
31			77	77	97	93	
32			68	85	80	28	
33			70	52	83	75	
34			63	58	58	83	
35			67	65	13	45	
36			62	70	27	73	
37			65	63	63	82	
38			73	78	87	83	
39			58	67	80	52	
40			57	85	78	55	

Table 13

Percentage of Correct Answers by Item on Standard Progressive Matrices for BN and CN Groups.

	SPM Subtest				
Item #	A	B	C	D	E
1	100	100	98	95	72
2	100	95	97	90	62
3	98	98	87	81	73
4	98	95	80	87	57
5	95	93	90	87	61
6	97	87	75	80	60
7	90	85	82	75	52
8	97	85	58	82	41
9	97	87	70	68	43
10	90	87	63	70	23
11	80	78	53	52	8
12	63	65	32	28	20

Table 14

Group Comparisons with Kruskal-Wallis for Mean Ranks of BCT and SPM Validity Indexes, and SPM Gross Errors.

	BC	BN	CN	NF	SF	d.f.	X^2
BCTBVI	59.8	73.3	55.6	100.2	88.6	4	30.57***
BCTTSVI	66.3	74.5	42.2	108.8	88.8	4	41.57***
BCTIVI	62.9	68.5	46.3	104.1	95.7	4	45.75***
BCTCVI	64.0	72.6	59.5	98.3	83.1	4	30.23***
BCTCPVI	60.9	71.0	54.5	97.5	93.7	4	31.24***
SPMVI	57.8	69.6	43.7	122.1	84.3	4	69.27***
SPMGE	66.1	62.9	27.8	118.2	104.5	4	81.41***

***p<.001

Table 15

Intercorrelations of Validity Indexes.

	BCTBVI	BCTTSVI	BCTIVI	BCTCVI	BCTCPVI	SPMVI
BCTBVI	1.000	.947	.937	.887	.861	.446
BCTTSVI		1.00	.915	.947	.908	.439
BCTIVI			1.00	.867	.827	.425
BCTCVI				1.00	.901	.516
BCTCPVI					1.00	.536
SPMVI						1.00

Note. Correlation coefficients reflect Pearson's r. N = 150. All correlations were significant at the $p < .001$ level.

Table 16

Group Comparisons with ANOVA or MANOVA for Mean Total Scores on BCT and SPM, BCT and SPM Validity Indexes, SPM Rate of Decay, and SPM Gross Errors.

	BC	BN	CN	NF	SF	d.f.	F[a]
BCT	69.27	67.07	42.20	94.30	75.63	4	12.01***
SPM	40.80	38.50	51.83	22.57	30.53	4	31.44***
BCTBVI	.33	1.07	.13	2.60	1.23	4	8.40***
BCTTSVI	1.40	2.50	.60	5.10	3.37	4	11.58***
BCTIVI	.57	1.03	.00	2.77	1.87	4	9.79***
BCTCVI	.01	.40	.00	1.13	.41	4	7.22***
BCTCPVI	.43	1.20	.13	2.47	1.67	4	6.74***
SPMVI	.57	.77	.00	4.90	1.90	4	29.65***
SPMRD[b]	16.33	16.03	8.20	10.10	13.67	6	9.73***
SPMGE[c]	5.53	5.50	.70	17.07	12.23	5	25.81***

***$p<.001$

[a] The F statistic was derived from MANOVA for the VI's.

[b] ANQ and IQ were covariates.

[c] ANQ was a covariate.

Table 17

Planned Comparisons for Validity Indexes.

Variable	Comparison	Value of Contrast	t	d.f.	p
BCTBVI	BC,BN,CN < NF,SF	4.22	4.160	60.08	.000
	BC,BN,CN < NF	6.27	3.683	33.17	.001
	BC,BN,CN < SF	2.17	2.381	46.13	.021
BCTTSVI	BC,BN,CN < NF,SF	8.10	5.285	80.11	.000
	BC,BN,CN < NF	10.80	4.758	35.16	.000
	BC,BN,CN < SF	5.40	2.988	39.33	.005
BCTIVI	BC,BN,CN < NF,SF	5.27	5.096	75.34	.000
	BC,BN,CN < NF	6.67	4.216	34.38	.000
	BC,BN,CN < SF	3.87	3.327	39.79	.002
BCTCVI	BC,BN,CN < NF,SF	1.88	3.659	54.69	.001
	BC,BN,CN < NF	2.93	3.322	32.16	.002
	BC,BN,CN < SF	0.83	1.867	43.08	.069
BCTCPVI	BC,BN,CN < NF,SF	4.43	4.106	69.25	.000
	BC,BN,CN < NF	5.63	3.254	34.32	.003
	BC,BN,CN < NF,SF	3.23	2.993	44.57	.004
SPMVI	BC,BN,CN < NF,SF	8.83	7.687	64.22	.000
	BC,BN,CN < NF	13.33	7.620	30.90	.000
	BC,BN,CN < SF	4.33	3.043	31.92	.005

Note. The t statistic does not assume equal variances. The p values are two-tailed.

Table 18

Post-Hoc Analyses of Group Differences on Validity Indexes.

	BC		BN		CN		SF	
	I-J	p	I-J	p	I-J	p	I-J	p
BCTBVI								
NF	2.27	.002	1.53	.127	2.47	.001	1.37	.132
SF	0.90	.026	0.17	.500	1.10	.002		
BCTTSVI								
NF	3.70	.000	2.60	.041	4.50	.000	1.73	.232
SF	1.97	.011	0.87	.486	2.77	.000		
BCTIVI								
NF	2.20	.001	1.73	.048	2.73	.000	0.90	.401
SF	1.30	.012	0.83	.374	1.83	.000		
BCTCVI								
NF	1.07	.005	0.73	.166	1.13	.003	0.70	.140
SF	.37	.063	.00	.500	.43	.014		
BCTCPVI								
NF	2.03	.007	1.27	.290	2.33	.001	0.80	.456
SF	1.23	.009	0.47	.498	1.53	.000		
SPMVI								
NF	4.33	.000	4.13	.000	4.87	.000	3.00	.001
SF	1.33	.063	1.13	.136	1.87	.002		

Note. I-J refers to the mean difference between scores for the groups listed in the column on the left (I) and the groups listed across the top (J). Homogeneity of variance assumption was not met, so Tamhane T2 was used for comparisons.

Table 19

X^2 Comparisons for BCTChance, SPMChance, and SPMDIS.

	BC	BN	CN	NF	SF	Total	X^2	d.f.	p
BCTChance									
Below Chance	0	0	0	1	0	1	4.07	4	.402
Above Chance	30	30	30	29	30	149			
SPMChance									
Below Chance	0	0	0	5	1	6	16.32	4	.003
Above Chance	30	30	30	25	29	144			
SPMDIS									
Discrepancy	5	5	1	15	11	37	43.65	8	.000
No Discrepancy	24	24	29	8	17	102			
Not Scorable	1	1	0	7	2	11			

Table 20

Planned Comparisons for SPMRD.

Comparison	F	d.f.	p
BC,BN,CN > NF,SF	0.35	1	.556
NF > SF	14.49	1	.001
BC > BN	0.01	1	.920

Note. Equal variances are assumed. ANQ and IQ were covariates.

Table 21

Post-Hoc Comparisons of Group Differences on SPMRD and SPMGE.

	BC		BN		CN		SF	
	I-J	p	I-J	p	I-J	p	I-J	p
SPMRD[a]								
NF	-6.23	.001	-5.93	.002	1.90	.400	-3.57	.214
SF	-2.67	.255	-2.37	.313	5.47	.005		
CN	-8.13	.000	-7.83	.000				
SPMGE[b]								
NF	11.53	.000	11.57	.000	16.37	.000	4.83	.153
SF	6.70	.003	6.73	.002	11.53	.000		
CN	-4.83	.003	-4.80	.001				

Note. I-J refers to the mean difference between scores for the groups listed in the column on the left (I) and the groups listed across the top (J).

[a]Homogeneity of variance assumption was met for SPMRD. Tukey HSD was used for comparisons.

[b]Homogeneity of variance assumption was not met for SPMGE. Tamhane T2 was used for comparisons.

Table 22

Planned Comparisons for SPMGE.

Comparison	F	d.f.	p
BC,BN,CN > NF,SF	62.09	1	.000
NF > SF	8.95	1	.003
BC > BN	0.001	1	.969

Note. Equal variances are not assumed. ANQ was a covariate.

Table 23

Discriminant Function Hit Rates for Clinical Versus Faking Groups.

Variable	Hit Rate[a]	Sensitivity Rate[b]	Specificity Rate[c]
TBCT	60.0%	63.3%	56.7%
TSPM	69.2%	73.3%	65.0%
BCTTSVI	65.8%	51.7%	80.0%
SPMVI	70.0%	48.3%	91.7%
SPMGE	75.0%	68.3%	81.7%
Overall	76.7%	66.7%	86.7%

Note. N = 120; Clinical is comprised of BC and BN, and faking is comprised of NF and SF.

[a]Hit Rate = % participants correctly categorized, N = 120.

[b]Sensitivity Rate = % naive and sophisticated fakers correctly categorized as malingering, N = 60.

[c]Specificity Rate = % clinical subjects correctly categorized as nonmalingering, N = 60.

Table 24

Discriminant Function Hit Rates for Nonfaking Versus Faking Groups.

Variable	Hit Rate	Sensitivity Rate	Specificity Rate
TBCT	65.3%	66.7%	64.4%
TSPM	78.0%	81.7	75.6%
BCTTSVI	72.7%	65.0%	77.8%
SPMVI	80.0%	65.0%	90.0%
SPMGE	80.0%	75.0%	83.3%
Overall	81.3%	73.3%	86.7%

Note. N = 150; CN is included in nonfaking group.

[a]Hit Rate = % participants correctly categorized, N = 150.

[b]Sensitivity Rate = % naive and sophisticated fakers correctly categorized as malingering (True Positives), N = 60.

[c]Specificity Rate = % clinical subjects and controls correctly categorized as nonmalingering (True Negatives), N = 90.

Table 25

Comparisons Between Clinical Groups on Demographic and Comparison Variables.

Variable	BC	BN	Difference
Gender	M = 16	M = 22	n.s.
	F =14	F = 8	
Age	42.5 (13.5)	39.6 (12.3)	n.s.
ANQ	20.7 (9.5)	16.7 (6.1)	n.s.
IQ	91.7 (10.8)	94.4 (17.1)	n.s.
CHI	15	25	$X^2_{(5)} = 11.11$, $p = .049$
Rey	13.03 (2.36)	11.60 (2.59)	$t_{(58)} = 2.24$, $p = .029$
Months Post-Injury	34.3 (37.2)	93.97 (83.8)	$t_{(40.02)} = -3.57$, $p = .001$
TBCT	69.3 (27.2)	67.1 (33.3)	n.s.
TSPM	40.8 (12.3)	38.5 (13.4)	n.s.
BCTBVI	0.33 (0.84)	1.07 (2.23)	n.s.
BCTTSVI	1.40 (1.35)	2.50 (3.42)	n.s.
BCTIVI	0.57 (1.10)	1.03 (2.25)	n.s.
BCTCVI	0.01 (0.25)	0.40 (1.07)	n.s.
BCTCPVI	0.43 (1.04)	1.20 (2.51)	n.s.
SPMVI	0.57 (1.25)	0.77 (1.14)	n.s.
SPMRD	16.33 (6.14)	16.03 (7.23)	n.s.
SPMGE	5.53 (6.39)	5.50 (5.83)	n.s.
BCTChance	Below = 0	Below = 0	n.s.
	Above = 30	Above = 30	

Table Continues

Table 25 Continued

SPMChance	Below = 0	Below = 0	n.s.
	Above = 30	Above = 30	
SPMDIS	0 = 5	0 = 5	n.s.
	1 = 24	1 = 24	
	2 = 1	2 = 1	

Note. Unless otherwise specified, mean values for each group are given, with standard deviations in parentheses.

Table 26

Comparisons Between Faking Groups on Demographic and Comparison Variables.

Variable	NF	SF	Difference
Gender	M = 9	M = 11	n.s.
	F=21	F = 19	
Age	32.8 (13.5)	29.3 (4.3)	n.s.
ANQ	6.1 (4.5)	3.1 (2.0)	$t_{(39.3)} = 3.17$, $p = .003$
IQ	102.6 (11.4)	122.2 (11.9)	$t_{(58)} = 6.50$, $p < .001$
TBCT	94.3 (28.4)	75.6 (31.2)	$t_{(58)} = 2.42$, $p = .018$
TSPM	22.6 (9.8)	30.5 (11.8)	$t_{(58)} = -3.03$, $p = .004$
BCTBVI	2.60 (3.00)	1.23 (1.45)	$t_{(41.9)} = 2.24$, $p = .030$
BCTTSVI	5.10 (3.94)	3.37 (3.040	n.s.
BCTIVI	2.77 (2.76)	1.87 (1.95)	n.s.
BCTCVI	1.13 (1.57)	0.41 (0.73)	$t_{(40.9)} = 2.22$, $p = .032$
BCTCPVI	2.47 (3.03)	1.67 (1.75)	n.s.
SPMVI	4.90 (3.14)	1.90 (2.54)	$t_{(55.5)} = 4.07$, $p < .001$
SPMRD	10.10 (7.07)	13.67 (6.40)	$t_{(58)} = -2.05$, $p = .045$
SPMGE	17.07 (7.61)	12.23 (7.54)	$t_{(58)} = 2.47$, $p = .016$
BCTChance	Below = 1	Below = 0	n.s.
	Above = 29	Above = 30	
SPMChance	Below = 5	Below = 1	n.s.
	Above = 25	Above = 29	

Table Continues

Table 26 Continued

SPMDIS (by score)	0 = 15	0 = 11	$X^2_{(4)} = 6.63$, $p = .036$
	1 = 8	1 = 17	
	2 = 7	2 = 2	

Note. Unless otherwise specified, mean values for each group are given, with standard deviation is parentheses.

Table 27

Hit, Sensitivity, and Specificity Rates for Detection Methods.

Detection Method	Hit Rate[a]	Sensitivity Rate[b]	Specificity Rate[c]
TBCT > 105	64.7%	21.7%	93.3%
TSPM < 22	72.0%	43.3%	85.5%
BCTBVI> 3[d]	66.7%	20.0%	97.8%
BCTTSVI > 5	70.7%	31.7%	96.7%
BCTIVI > 3	66.0%	27.0%	95.5%
BCTCVI > 1	66.0%	19.3%	96.7%
BCTCPVI > 3	66.7%	22.3%	95.5%
SPMVI > 2	76.0%	48.3%	94.4%
BCT Below Chance	40.7%	1.7%	100%
SPM Below Chance	64.0%	10.0%	100%
SPMRD[e]	66.0%	31.7%	88.9%
SPMGE > 2	78.0%	56.7%	92.2%
SPMDIS =1	77.2%	43.3%	85.5%

[a]Hit Rate = % participants correctly categorized, N = 150.

[b]Sensitivity Rate = % naive and sophisticated fakers correctly categorized as malingering (True Positives), N = 60.

[c]Specificity Rate = % clinical subjects and controls correctly categorized as nonmalingering (True Negatives), N = 90.

[d]Bolter et al. (1992) suggested a slightly lower cut-off, which was less than optimal for distinguishing fakers in this study.

[e]Cut-off scores varied with Total SPM score.

Table 28

Cut-Off Scores for SPMRD Based on Total SPM Score for the Present and Prior Studies.

	Total SPM Score[a]									
	10	15	20	25	30	35	40	45	50	55+
Present Study[b]										
Avg RD		19	14.5	17.3	24.3	22	18	17	10.6	4.3
# Cases	0	2	2	3	4	4	3	8	14	20
Cut-off		10	7	9	12	11	9	9	5	2
Prior Study[c]										
Avg RD	13.5	17.5	20.5	22	21	17	15.5	13.5	12	4.5
Cut-off	7	9	10	11	12	10	8	7	6	2

[a]Subject scores were placed into one of the 10 total score groups that most closely corresponded to their scores.

[b]The BN and CN groups were used to derive the cut-off scores used for comparisons in the present study. N = 60.

[c]Gudjonsson & Shackleton, 1986; Avg RD reflects the average between the scores for their sample and those found in Raven's (1960) manual.

Table 29

Differential Sensitivity Rates for Naive Versus Sophisticated Fakers.

Detection Method	NF	SF	Total
TBCT > 105	26.7%	16.7%	21.7%
TSPM < 22	63.3%	23.3%	43.3%
BCTBVI > 3[c]	33.3%	6.7%	20.0%
BCTTSVI > 5	40.0%	23.3%	31.7%
BCTIVI > 3	36.7%	16.7%	27.0%
BCTCVI > 1	30.0%	6.7%	19.3%
BCTCPVI > 3	30.7%	16.7%	22.3%
SPMVI > 2	66.7%	30.0%	48.3%
BCT Below Chance	3.3%	0.0%	1.7%
SPM Below Chance	16.7%	3.3%	10.0%
SPMRD[d]	46.7%	16.7%	31.7%
SPMGE > 2	66.7%	46.7%	56.7%
SPMDIS =1	50.0%	36.7%	43.3%

[a]Hit Rate = % subjects correctly categorized, N = 150.

[b]Sensitivity Rate = % naive and sophisticated fakers correctly categorized as malingering (True Positives), N = 60.

[c]Bolter et al. (1992) suggested a slightly lower cut-off, which was less than optimal for distinguishing fakers in this study.

[d]Cut-off scores varied with Total SPM score.

Table 30

Discriminant Function Hit Rates for BCT Versus SPM.

Comparison Groups	Variable	Hit Rate[a]	Sensitivity Rate[b]	Specificity Rate[c]
Clinical vs. Faking	BCT	68.3%	60.0%	76.7%
N = 120	SPM	76.7%	66.7%	86.7%
Nonfaking vs. Faking	BCT	71.3%	63.3%	76.7%
N = 150	SPM	80.7%	76.7%	83.3%

Note. BCT detection methods included BCTTSVI and TBCT. SPM detection methods included SPMVI, SPMGE, and TSPM.

[a]Hit Rate = % participants correctly categorized.

[b]Sensitivity Rate = % naive and sophisticated fakers correctly categorized as malingering (True Positives).

[c]Specificity Rate = % subjects correctly categorized as nonmalingering (True Negatives).

REFERENCES

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

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. American Psychologist, 47, 1597-1611.

Aubrey, J. B., Dobbs, A. R., & Rule B. G. (1989). Laypersons' knowledge about the sequelae of minor head injury and whiplash. Journal of Neurology, Neurosurgery, and Psychiatry, 52, 842-846.

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

Beaumont, J. G. (1975). The validity of the Category Test administered by on-line computer. Journal of Clinical Psychology, 31, 458-462.

Bender, L. (1938). A visual motor Gestalt test and its clinical use. American Orthopsychiatric Association, Research Monographs, No. 3.

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

Benton, A. L., & Spreen, O. (1961). Visual memory test: The simulation of mental incompetence. Archives of General Psychiatry, 4, 79-83.

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

Bernard, L. C., McGrath, M.J., & 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, T. R. D., Lamb, D. G., Wetter, M. W., Baer, R. A., & Widiger, T. A. (1994). Ethical considerations in research on coached malingering. Psychological Assessment, 6, 16-17.

Bigler, E. D. (1990). Neuropsychology and malingering: Comment on Faust, Hart and Guilmette (1988). Journal of Consulting and Clinical Psychology, 58, 244-247.

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

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. Journal of Clinical and Experimental Neuropsychology, 15, 170-182.

Binder, L. M., & 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, 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, Silas B. Hays Army Community Hospital, Department of Psychiatry, Fort Ord, CA.

Bolter, J. F., Picano, J. J., & Zych, K. (1985, October). Item error frequencies on the Halstead Category Test: An index of performance validity. Paper presented at the meeting of the National Academy of Neuropsychologists, Philadelphia, PA.

Brandt, J. (1988). Malingered amnesia. In R. Rogers (Ed.).Clinical assessment of malingering and deception (pp. 65-83). New York: Guilford Press.

Brown, L., Sherbenou, R. J., & Johnson, S. L. (1982). Test of Nonverbal Intelligence: A language-free measure of cognitive ability. Austin, TX: Pro-Ed.

Brownlee, K. A. (1965). Statistical theory and methodology in science and engineering (2nd ed.). New York: John Wiley.

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

Butcher, J., Dahlstrom, W., Graham, J., Tellegen, A., & Kaemmer, B. (1989). Manual for administering and scoring the MMPI-2. Minneapolis: University of Minnesota Press.

Charter, R. A. (1994). Determining random responding for the Category, Speech-Sounds Perception, and Seashore Rhythm Tests. Journal of Clinical and Experimental Neuropsychology, 16, 744-748.

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

Davidson, H., Suffield, B., Orenczuk, S., Nantau, K., & Mandel, A. (1991, February). Screening for malingering using the memory for fifteen items test (MFIT). Paper presented at the International Neuropsychological Society annual convention, San Antonio, TX.

DeFilippis, N. A., McCampbell, E. (1979). The Booklet Category Test: Research & clinical form. Manual. Odessa, FL: Psychological Assessment Resources.

DeFilippis, N. A., McCampbell, E., & Rogers, P. (1979). Development of a booklet form of the Category Test: Normative and validity data. Journal of Clinical Neuropsychology, 1, 339-342.

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

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

Faust, D., Guilmette, T. J., Hart, K., Arkes, H. R., Fishburne, F. J., & Davey, L. (1988). Neuropsychologists' experience and judgement accuracy. Archives of Clinical Neuropsychology, 3, 145-163.

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, 56 578-582.

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

Fee, C. R., & Rutherford, W. H. (1988). A study of the effect of legal settlement on post-concussion syndrome. Archives of Emergency Medicine, 5, 12-17.

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

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

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

Golden, C. J., Hammeke, T. A., & Purisch, A. D. (1980). The Luria-Nebraska Neuropsychological Battery: Manual. Los Angeles: Western Psychological Services.

Golden, C. J., Zillmer, E., & Spiers, M. (1992). Neuropsychological assessment and intervention. Springfield Ill.: U.S.A.

Gorman, W. F. (1982). Defining malingering. Journal of Forensic Sciences, 27, 401-407.

Gouvier, W. D., Uddo-Crane, M., & Brown, L. M. (1988). Base rates of post-concussional symptoms, Archives of Clinical Neuropsychology, 3, 273-278.

Green, K. E., & Kluever, R. C. (1992). Components of item difficulty of Raven's Matrices. The Journal of General Psychology, 119, 189-199.

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

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

Guilmette, T., Hart, K., & Giuliano, A. (1993). Malingering detection: The use of a forced choice method in identifying organic versus simulated memory impairment. The Clinical Neuropsychologist, 7, 59-69.

Guilmette, T. J., Sparadeo, F. R., Whelihan, W., & Buongiorno, G. (1994). Validity of neuropsychological test results in disability evaluations. Perceptual and Motor Skills, 78, 1179-1186.

Haines, M. E., & Norris, M. P. (1995). Detecting the malingering of cognitive deficits: An update. Neuropsychology Review, 5, 125-148.

Hall, H. V., & Pritchard, D. A. (1997). Detecting malingering and deception. Delray Beach, FL: St. Lucie Press.

Halstead, W. C. (1947). Brain and intelligence. Chicago, IL: The University of Chicago Press.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. Journal of Psychology, 10, 249-254.

Heaton, R. K. (1981). A manual for the Wisconsin Card Sorting Test. Odessa, FL: Psychological Assessment Resources.

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

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

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, 667-676.

Jarvis P. E., & Barth, J. T. (1984). Halstead-Reitan Test Battery: An interpretive guide. Odessa, FL: Psychological Assessment Resources.

Jensen, J. A., & Armstrong, R. J. (1985). Slosson Intelligence Test (SIT) for children and adults: Expanded norms tables application and development. East Aurora, NY: Slosson Educational Publications.

Kaufman, A. S. (1990). Assessing adolescent and adult intelligence. Boston: Allyn and Bacon, Inc.

Klett, W. G., Watson, C. G., & Hoffman, P. T. (1986). The Henmon-Nelson and Slosson tests as predictors of WAIS-R IQ. Journal of Clinical Psychology, 42, 343-347.

Lees-Haley, P.R., & Brown, R. S. (1993). Neuropsychological complaint base rates of 170 personal injury claimants. Archives of Clinical Neuropsychology, 8, 203-209.

Lezak, M. (1995). Neuropsychological Assessment (3rd ed.). New York: Oxford.

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

Martin, R. C., Hayes, J. S., & Gouvier W. D. (1996). Differential vulnerability between postconcussion self-report and objective malingering tests in identifying simulated mild head injury. Journal of Clinical and Experimental Neuropsychology, 18, 265-275.

McCaffrey, R. J., & Lynch, J. K. (1992). A methodological review of "Method Skeptic" reports. Neuropsychology Review, 3, 235-248.

Melendez, F. (1978). Revised manual for the Adult Neuropsychological Questionnaire. Odessa, FL: Psychological Assessment Resources.

Mendelson, G. (1985). Compensation neurosis: An invalid diagnosis. The Medical Journal of Australia, 142, 561-564.

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

Mercer, W N. (1994). Performance of brain-injured verses non-brain-injured individuals on three versions of the Category Test. Unpublished Doctoral Dissertation, University of North Texas, Denton.

Milanovich, J. R., Axelrod, B. N., & Millis, S. R. (1995). Validation of the Simulation Index-Revised with a mixed clinical population. Archives of Clinical Neuropsychology, 11, 53-59.

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

Millis, S. R., & Kler, S. (1995). Limitations of the Rey fifteen-item test in the detection of malingering. The Clinical Neuropsychologist, 9, 241-244.

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

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

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

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

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

Pankratz, L., & Binder, L. M. (1997). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.). Clinical assessment of malingering and deception (pp. 223-236). New York: Guilford.

Paul, D. S., Franzer, M. D., Cohen, S. H., & Fremouw, W. (1992). An investigation into the reliability and validity of two tests used in the detection of dissimulation. International Journal of Clinical Neuropsychology, 4, 1-9.

Prigatano, G. P., & Amin, K. (1993). Digit Memory Test: Unequivocal cerebral dysfunction and suspected malingering. Journal of Clinical and Experimental Neuropsychology, 15, 537-546.

Raven, J. C. (1960). Guide to the SPM. London: H. K. Lewis.

Raven, J. C., Court, J. H., & Raven, J. (1977). Manual for Raven's Progressive Matrices and Vocabulary Scales. London: H. K. Lewis.

Rawling, P. J. (1992). The simulation index: A reliability study. Brain Injury, 6, 381-383.

Rawling, P. J. (1993). Simulation Index - Revised Manual. Balgowlah, Australia: Rawling & Associates.

Rawling, P., & Brooks, N. (1990). Simulation index: A method for detecting factitious errors on the WAIS-R and WMS. Neuropsychology, 4, 223-238.

Reitan, R. M., & Davison, L. A. (1974). Clinical neuropsychology: current status and applications. Washington, DC: W. H. Winston & Sons.

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

Reitan, R. M., & Wolfson, D. (1993). The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation (2nd ed.). Tucson, AZ: Neuropsychology Press.

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

Rey, A. (1964).L'Examen psychologique dans le cas d'encephalopathie traumatique. Archives de Psychologie, 37, 126-139.

Rogers, R. (1988). Researching dissimulation. In R. Rogers (Ed.) Clinical assessment of malingering and deception (pp. 309-327). New York: Guilford.

Rogers, R. (1990a). Development of a new classificatory model of malingering. Bulletin of the American Academy of Psychiatry and Law, 18, 323-333.

Rogers, R. (1990b). Models of feigned mental illness. Professional Psychology, 21, 182-188.

Rogers, R. (Ed.). (1997). Clinical assessment of malingering and deception (2nd ed.). New York: Guilford Press.

Rogers, R., Gillis, J. R., Bagby, R. M., & Monteiro, E. (1991). Detection of malingering on the structured interview of reported symptoms (SIRS): A study of coached and uncoached simulators. Psychological Assessment A Journal of Consulting and Clinical Psychology, 3, 673-677.

Rogers, R., Gillis, J. R., Dickens, S., & Bagby, R. M. (1991). Standardized assessment of malingering: Validation of the structured interview of reported symptoms. Psychological Assessment A Journal of Consulting and Clinical Psychology, 3, 89-96.

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

Rose, F. E., Hall, S., & Szalda-Petree, A. D. (1995). Portland Digit Recognition Test-computerized: Measuring response latency improves the detection of malingering. The Clinical Neuropsychologist, 9, 124-134.

Rosenhan, D. L. (1973). On being sane in insane places. Science, 179, 250-258.

Rosenfeld, J. P., Sweet, J. J., Chuang, J., Ellwanger, J., & Song, L. (1996). Detection of simulated malingering using forced choice recognition enhanced with event related potential recording. The Clinical Neuropsychologist, 10, 163-179.

Rothke, S. (1986). The role of set shifting cues on the Wisconsin Card Sorting Test and Halstead Category Test. The International Journal of Clinical Neuropsychology, 8, 11-14.

Rudell, A. P. (1991). The recognition potential contrasted with the P300. International Journal of Neuroscience, 60, 85-111.

Ruff, R. M., Wylie, T., & Tennant, W. (1993). Malingering and malingering-like aspects of mild closed head injury. Journal of Head Trauma Rehabilitation, 8, 60-73.

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, 3, 667-672.

Sivan, A. B. (1992). Benton Visual Retention Test (5th ed.). San Antonio, TX: The Psychological Corporation.

Slosson, R. L. (1982). Slosson Intelligence Test (2nd ed.). East Aurora, NY: Slosson Educational Publications.

Smith, A. (1975). Neuropsychological testing in neurological disorders. In W. J. Freidlander (Ed.), Advances in neurology (Vol. 7). New York: Raven Press.

Snedecor, G. W., & Cochran, W. G. (1967). Statistical methods (6th ed.). Ames, IA: Iowa State University Press.

Tenhula, W. N., & Sweet, J. J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. The Clinical Neuropsychologist, 10, 104-116.

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

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

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

Verlager, R., & Berg, P. (1991). The waltzing oddball. Psychophysiology, 28, 468-477.

Warrington, E. K. (1984). Recognition Memory Test. Berkshire, U. K.: NFER-Nelson.

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

Wechsler, D. (1974). Wechsler Memory Scale manual. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale-Revised (WAIS-R). San Antonio, TX: The Psychological Corporation.

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

Youngjohn, J. R., Burrows, L., & Erdal, K. (1995). Brain damage or compensation neurosis? The controversial Post-Concussion Syndrome. The Clinical Neuropsychologist, 9, 112-123.