This is to certify that the

dissertation entitled

DETECTION OF DECEPTION IN NEUROPSYCHOLOGICAL

TESTING: A MULTI-METHOD APPROACH

presented by

David Dean Cordry

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Mona Abeles
Major professor

Date March 25, 2002

DETECTION OF DECEPTION IN NEUROPSYCHOLOGICAL

TESTING: A MULTI-METHOD APPROACH

By

David Dean Cordry

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

2002

ABSTRACT

DETECTION OF DECEPTION IN NEUROPSYCHOLOGICAL TESTING: A MULTI-METHOD APPROACH

By

David Dean Cordry

An experiment to determine the utility of incorporating several methods to detect simulated malingering of cognitive problems was conducted. Patterns of performance, on neuropsychological tests, specifically a magnitude of error strategy, as well as responses on a self-report measure and nonverbal behaviors during a clinical interview were examined. College students were assigned to either a control or an experimental group where they were asked to simulate having a head injury for the purpose of attempting to win a $5 reward. When all variables were used to predict group membership, the discriminant function achieved sensitivity, or a true positive rate for malingering of 92% and a specificity, or true negative rate for controls of 100%, with an overall classification accuracy of 95%. Results are discussed within the context of the benefits of using a multi-method approach to detect feigning.

To those who hear the sound of one hand clapping.

ACKNOWLEDGMENTS

I would like to thank many people who made this particular pursuit of erudition possible. I would like to express my gratitude for my mentor and chairperson, Professor Abeles for all of his guidance and support in writing of this dissertation, as well as his advocacy on my behalf. Furthermore, I would like to thank Professor Levine for all of his assistance with the statistical analyses and insightful comments into the design of this research. I would also like to express my gratitude to Professors Winder and Thornton who have both been excellent clinical supervisors in addition to providing comments and criticisms that helped guide this work.

I would like to especially thank Angela McBride for all of her assistance in this project, not to mention all of the hours of Monopoly Junior that kept us both sane. Camilla Williams is also deserving of thanks for her assistance in this project. Several friends and colleagues helped make this possible, providing support and friendship that sustained me through this effort.

I would like to thank my family for providing me with the means to pursue this opportunity. Their emotional support and encouragement throughout my life has made many things possible for me.

I would like to thank my wife, Meg for all of her love and support through this endeavor. Without her by my side, none of my accomplishments would have been possible. Finally, I would like to thank my son, Maxim for coming into my life and giving me the motivation to complete this task, showing me what things truly matter, and filling my life with more joy than I would have thought possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Detection of Deception in Neuropsychological Testing: A Multi-method Approach

INTRODUCTION

Deception is used to some degree in all forms of social interactions. The various motivations for successfully deceiving another person include maintaining appropriate politeness, gaining another's approval, and attempts to deceive in order to achieve some financial reward, to name a few. The problem of deception also exists in the area of neuropsychological testing. Psychologists rely primarily on self-disclosure from their clients in order to make an appropriate diagnosis and begin the course of treatment. However, when psychologists are asked to evaluate clients to answer questions of impairment that have financial or legal incentives for their clients, naïve assumptions about the genuineness of their clients' efforts may reflect a form of self-deception.

The increasing use of neuropsychological test data to determine neuropsychological impairment in a legal setting may influence a client's approach to the testing procedure in order to achieve some type of financial reward (Lees-Haley, 1990). The possibility of contamination of test results by monetary or other incentives has lead to an increase in development of procedures to detect attempts to malinger—that is, deliberate simulation or exaggeration of a disability—with varying degrees of success (Lees-Haley and Fox, 1990; Bernard, 1991; Faust, 1995).

When one considers a diagnosis of malingering, it is important to note that malingering is not a dichotomous construct (Sweet, 1999). Individuals may perform in a manner consistent with malingering on some tasks, while performing at their actual ability level on others. It is important to consider that performance on any particular neuropsychological test can not serve as conclusive evidence of malingering, just as performance on some tests that is not consistent with malingering does not rule out malingering on other measures. Furthermore, estimates of the base rate of neuropsychological malingering in actual clinical populations have ranged from 7.5-15% (Trueblood & Schmidt, 1993) to 18-33% (Binder, 1993) suggesting that diagnosis of malingering occurs infrequently.

The aim of the current investigation is to evaluate the strategies used by students instructed to malinger. Specifically, do analog malingerers produce a different pattern of performance when attempting to simulate a head injury, than do students asked to perform to the best of their ability, or the performance of actual head injured patients demonstrated in previous research?

Detection of Malingering: Patterns of Performance on Neuropsychological Tests

One approach to the detection of exaggerated performance on neuropsychological tests involves the analysis of differences in performance on commonly used tests between malingerers, controls and impaired subjects. Being able to explain the procedures for distinguishing between a malingered performance and an impaired performance is a task likely to be performed by the clinical neuropsychologist during testimony in a litigated case (Iverson, Franzen & McCracken, 1991). The use of analyzing performances on existing and commonly used tests to differentiate a malingered performance represents an advantage over adding additional tests and procedures to an already cumbersome neuropsychological test battery (Bernard, McGrath, & Houston, 1993).

The majority of the research literature that has been done regarding the detection of malingering on neuropsychological tests has focused on exaggeration of memory deficits, rather than on faking of specific disorders such as learning disabilities, or focal brain lesions (Nies & Sweet, 1994). Memory complaints are among the most common complaints to be investigated by clinical neuropsychologists, and are an area for additional investigation in regard to ability to detect malingered performances (Bernard, 1990). Malingered memory deficits and complaints may occur more often in clients motivated to receive some type of compensation because they are presumably easy to demonstrate, yet hard to detect as exaggerated (Nies & Sweet).

A study by Brandt, Rubinsky and Lassen (1985) found that college students given instructions to fake a memory deficit did not differ in their performance on a free recall task from a head trauma group or a group of Huntington's patients; although they performed significantly below a control group. However, on a recognition trial, only the malingerers scored below chance levels on a forced-choice task. Although the authors of this study suggested that recognition might distinguish malingerers from normals or clinical groups, they cautioned that a substantial portion of naïve malingerers performed at or above chance levels. A performance on a forced-choice task below chance levels is suggestive of malingering according to binomial theory; that is, if a subject guesses randomly on a task where one must choose between two alternatives approximately 50% accuracy should be achieved (Guilmette, Hart, Giuliano, & Leininger, 1994).

However, since malingering represents a conscious attempt to simulate impairment, a malingering subject would be expected to monitor performance and choose the incorrect response when they know the correct one and thus perform below the level of chance (Iverson, Franzen, & McCracken, 1994).

Iverson, Franzen and McCracken (1991) used a 21-item word list to examine patterns of performance exhibited by college students instructed to fake memory impairment as compared to patterns demonstrated by patients with documented memory impairment. The investigators evaluated the performance of all participants on both a free recall task, as well as on a forced-choice recognition task, where subjects were given word pairs and asked to decide between the word from the original list and a distractor. They found that although there was no significant difference between the free recall performance of the malingerers and the memory-impaired participants, on the recognition task malingerers performed significantly more poorly than the memory-impaired participants. Furthermore, when cutoff points were selected in order to classify the group membership of participants, they were able to correctly identify 100% of malingerers using a cut off score that fell in the confidence level of chance performance on the recognition task, while only misidentifying one memory impaired participant. When more conservative cutoff points were chosen, the false positive rate was reduced to 0%, although the identification rate for malingerers dropped to 65%.

The pattern of performance demonstrated by participants instructed to fake memory deficits on forced-choice recognition measures may be indicative of their level of motivation to appear impaired. Binder, Villanueva, Howieson, and Moore (1993) examined the patterns of performance of patients with mild head trauma with differing levels of financial incentive on a recognition memory task. In their study, the investigators divided a group of patients with a history of mild head injury into a group that was seeking financial compensation for their symptoms and a group with definite evidence of brain dysfunction that were not seeking financial compensation. The group seeking financial compensation was further divided based on their performance on a task measuring motivation (Portland Digit Recognition Test; PDRT). Binder and his colleagues found that the mild head trauma group with low motivation, as measured by poor PDRT performance, demonstrated significantly worse scores on the recognition trial of the Rey Auditory Verbal Learning Test

(RAVLT) than did any other group. Their results suggest that analysis of performance on recognition measures may be useful in discriminating groups of patients who are exaggerating deficits and patients with actual memory impairment.

Bernard (1990) found that using stepwise discriminant function analyses, he was able to achieve a 77% accuracy rate in predicting participants group membership as either a control or a simulated malingerer. His study found that performance on the recognition trial of the RAVLT and free recall performance on the Complex Figure Task (CFT) were the primary indicators in his analyses that detected malingering participants. Using the discriminant function from this study, Bernard, Houston, and Natoli (1993) achieved an overall accuracy rate of 86% in their subsequent study. Furthermore, they did not misidentify any control subjects as malingerers and had a 75% true positive rate in identifying malingerers.

Bernard (1991) extended his analysis of patterns of performance on neuropsychological tests to detect malingering to include an analysis of the serial position effects during free recall on the RAVLT. He hypothesized that malingerers might produce a characteristic pattern of free recall in their attempts to exaggerate memory deficits that would discriminate them from controls. Bernard found that a group of college students given instructions to simulate a memory deficit recalled significantly fewer words from the first third of the word list than from the last third. Bernard interpreted the absence of a U-shaped recall curve, which was demonstrated by controls, as indicative of the intentional suppression of items from the first third of the word list by malingering participants. Furthermore, subjects with closed head injuries, having documented brain injury demonstrated a recall pattern opposite of the malingerers; that is, they demonstrated a significantly greater level of recall from the first third of the list than from the last third. This finding suggests that analysis of the pattern of free recall may be of use in differentiating malingerers from patients with actual neuropsychological deficits.

The use of the serial position of free recall has only been reported in two literature studies and has had inconsistent support in its use for distinguishing between malingerers and control participants. Bernard, Houston and Natoli (1993) found that simulated malingerers demonstrated a U-shaped curve on a free recall task similar to that of controls, suggesting that the use of serial position patterns of performance on free recall may not be a solid indicator of exaggerated deficits. Likewise, Iverson, Franzen and McCracken (1991) were not able to differentiate between malingerers, controls or memory disordered patients on the basis of serial position curves produced during free recall on a verbal memory task; however, their research protocol did not involve repeated presentation of a word list. These findings suggest that their exists a considerable amount of variability in the performance of participants asked to malinger.

Furthermore, Wiggins and Brandt (1988) found that participants instructed to malinger displayed serial position curves similar to normals on a verbal memory task. Wiggins and Brandt asked their participants to respond to test items as though they had memory problems, but did not include any simulation of secondary gains. Interestingly, they did find that the malingering group performed significantly below a group of memory-disordered patients on a recognition measure but not on a free-recall measure.

Many studies of malingering of memory deficits have relied on the analysis of one memory measure. However, in clinical practice several measures of memory may be administered in addition to tests specifically designed to detect malingering. Iverson and Franzen (1996) utilized a battery of five objective assessment measures to enhance their accuracy in detecting experimental malingerers. They established cutoff scores for which there were no false positives and compared these scores for their individual accuracy in correctly classifying participants. Similar to previous studies, they found the forced-choice recognition task was effective in correctly identifying 69% of experimental malingerers. By combining all of the cutoff scores from the memory measures used, and using deficient performance on any one of the measures used, resulted in the most accurate classification rate of 92.5%. This finding supports the use of a battery of tests to detect malingered performances.

Trueblood and Schmidt (1993) used a below chance performance on a forced-choice test (Symptom Validity Test; SVT) to detect malingerers from a group of patients referred for neuropsychological evaluation. In addition they identified patients who demonstrated questionable validity on neuropsychological procedures. These researchers matched participants with malingered and questionable performances with controls and used a variety of neuropsychological procedures to discriminate between malingered performances and performances of head injured patients. They established cutoff scores that discriminated malingerers while minimizing false positives, and found that the California Verbal Learning Test *(CVLT)* Recognition score was one of the measures that performed well as a validity indicator, identifying 9 of 16 probable malingerers. They found that they were able to identify 13 of 16 probable malingerers when they used several measures and found more than two test performances falling below the malingering cutoff.

The utility of the CVLT in identifying malingering in participants with more specific instructions on faking amnesia has been also been demonstrated (Coleman, Rapport, Millis, Ricker, & Farchione, 1998). Coleman and her colleagues found that analog malingerers who were coached on how to fake memory problems, performed similarly to head injured patients on free-recall indices; however, on more subtle measures of learning such as recognition and slope of learning, they performed similarly to naïve malingerers. Coleman et al. used logistic regression and were able to achieve a specificity rate of 85.2% and sensitivity (to malingering) of 90.5% by including recognition discriminability in their analyses. The CVLT has also been found to be useful in discriminating between minor head injury patients with poor motivation from patients with severe head injuries (Millis, Putnam, Adams & Ricker, 1995). Discriminant function analysis of several CVLT variables correctly classified 91% of the patients with minor head injuries according to Millis and his colleagues. They found that the CVLT demonstrated both a high degree of sensitivity in detecting poor motivation and a high degree of specificity in correctly identifying patients with documented severe head injuries.

Another strategy for detecting malingering involves the analysis of magnitude of errors. Martin, Franzen, and Orey (1998) had 10 independent raters assign a probability of selection value to each answer choice on a multiple-choice recognition task, which followed the presentation of a story that participants were told to remember (Logical Memory; Wecshler Memory Scale-Revised). Martin and his colleagues averaged the selection probability values and were able to achieve a .85 inter-rater agreement. They found that students given instructions to fake memory impairment and patients who were identified as having questionable motivation were significantly more likely to select low probability responses than either controls or patients with documented closed-head injuries. The incorporation of a magnitude of error approach provided additional classification accuracy beyond that found when just the total recognition memory score was used. Furthermore, certain items were only selected by the malingering and questionable motivation group, suggesting that the magnitude of error approach may result in "critical items" that provide additional indications of exaggerated memory impairment. This strategy is also advantageous in that it can potentially be applied to neuropsychological tests that are commonly used during assessment, once probability values are assigned.

Personality Measures in the Detection of Neuropsychological Malingering

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemer, 1989) is one of the most widely used measures in the course of neuropsychological assessments (Lezak, 1995). Assessment of emotional adjustment and personality is a routine part of neuropsychological evaluations. Furthermore, the MMPI-2 consists of a number of items that describe neurological symptoms that are associated with head injury. When head injuries are assessed as part of a litigation process, the forensic nature of these evaluations raises the possibility of response distortion and exaggeration on all measures, including the MMPI-2. One advantage of using the MMPI-2 in a forensic context is the extensive body of research on the validity scales included in the measure.

Malingering of symptoms for the purpose of secondary gain can occur as either deliberately poor performance on neuropsychological measures, exaggeration of somatic complaints, or as a combination of both (Larrabee, 1998). Although no specific closed head injury profile has been established, certain trends of performance have been demonstrated (Berry et al., 1995). Generally, patients with closed head injuries produce elevated scores on scales F, 1, 2, 3, 7 and 8, although these scores are not always elevated to clinical levels (Alfano, Neilson, Paniak, & Finlayson, 1992; Gass & Russell, 1991; Diamond, Barth & Zillmer, 1988; Gass, 1991; Beery et al., 1995). The specificity of MMPI-2 profiles for closed head injured patients improves when the presence of litigation and other factors influencing motivation are taken into account and specific dissimulation scales are used (Larrabee, 1998).

Several MMPI -2 scales have demonstrated utility in detecting dissimulation of psychiatric symptoms, and have been suggested as being of value in detecting exaggeration of deficits in a neuropsychological context. The F (Frequency) scale is a validity scale on the MMPI-2 that consists of answers that are infrequently endorsed in the

scored direction by either disturbed individuals or normals (Lees-Haley, 1991). Persons who demonstrate elevated F scores typically admit to a wide range of psychological problems. The Fb scale (back F; Butcher et al., 1989) was constructed in a similar manner as the F scale, only using items taken from the second half of the MMPI-2. The Dissimulation Index, or F – K (Gough, 1950) was also developed as a method for detecting malingering of psychiatric syndromes. Obvious and Subtle subscales (O – S; Wiener, 1948) were also developed as a means of detecting malingering based on the number of items commonly known to indicate psychopathology versus the items that represent more subtle indications of pathology. The Dissimulation Scale (Ds; Gough, 1954) was originally constructed of 74 items from the original MMPI, thought to assess inaccurate stereotypes of neurotic psychopathology, and has been updated using 58 items on the MMPI-2 (Ds2; Berry, et al., 1995). Another validity scale that has been described by Arbisi and Ben-Porath (1993) as the F – Psychopathology (Fp) scale consists of 27 items rarely endorsed by psychiatric patients. Lees-Haley, English and Glenn (1991) presented an additional scale, the Fake Bad Scale (FBS) specifically designed to detect malingering among personal injury litigants.

In a comparison of MMPI-2 scores of personal injury claimants who were suspected of malingering and litigants not suspected of malingering, Lees-Haley (1991) found that suspected malingerers produced both higher F scores and higher F - K scores. However, the utility of the MMPI - 2 for detecting malingering in neuropsychological testing, was not specifically addressed in this study, as patients with other types of claims, such as Post-traumatic Stress Disorder were also included in the sample.

Analog malingerers have been found to produce similar overall levels of deficits, although their performance patterns are distinguishable (Heaton, Smith, Lehman, & Vogt, 1978). Heaton and his colleagues found that participants instructed to fake a head injury produced significantly higher scores on the F scale as well as on clinical scales 1, 3, 6, 7, and 8 than non-litigating head-injured participants. Furthermore, they found that head-injured patients and the malingering participants performed poorly on different neuropsychological measures. Heaton and his colleagues were able to correctly classify 100% of their participants using discriminant function analysis of the neuropsychological tests alone. In a separate discriminant function analysis using the MMPI scores, 94% of their participants were correctly classified, suggesting

that the MMPI may be of use in detecting malingering when used in conjunction with other neuropsychological measures.

Berry et al. (1995) compared MMPI -2 validity scales of normal controls, participants asked to feign head injury and given a description of closed head injury symptoms, litigating head injury patients, and non-litigating head injury patients. Berry et al., used the F, F - K, Fb, Ds and Fp scales in their analyses and found that the analog malingerers were distinguishable from the normal controls on all the included scales. Furthermore, litigating head injury patients were distinguished from non-litigating patients on the validity scales, and produced a similar pattern of elevated scores on the clinical scales as the analog malingerers.

The presence of elevations on clinical scales 1 and 3 are often seen in neuropsychological settings because the item content of these scales reflects somatic symptoms of pain, paresthesias, and malaise (Gass, 1991). However, only one F scale item is included on these scales, which has led some investigators to question reliance on the F scale to detect malingered performances (Larrabee, 1998). The FBS was designed to detect a mixture of faking good and faking bad response styles seen in personal injury claimants (Lees-Haley, English, & Glenn, 1991). Lees-Haley et al. hypothesized this combination of response styles, as representing several goals of a malingerer. These goals include an attempt by malingering persons to appear honest, appear psychologically normal except as influenced by the injury, and avoid admitting pre-existing psychopathology, while presenting a degree of complaints that suggests a believable degree of disability.

Larrabee (1998) found that the standard validity scales F, F- K, Fb, and Fp were insensitive to detection of exaggerated somatic complaints in patients who performed below suggested cutoff scores of neuropsychological tests of malingering. In addition to an elevated FBS scale, Larrabee found that medically and neurologically normal litigants demonstrated elevations on scale 1 and 3 greater than a group of chronic pain patients and a group of head injured patients. The FBS demonstrated good sensitivity, detecting 91% of patients considered to be malingering based on other test performances.

Millis, Putnam and Adams (1995) reported similar findings in their investigation of mild head injury patients involved in litigation, scoring below chance, indicating probable malingering on a neuropsychological test. Although the litigants in their study produced significant elevations over patients with moderate to severe head injuries on validity scales F, Fb, and Fp, *and* FBS, the FBS was found to have the best diagnostic efficiency. Furthermore, significant elevations on the clinical scales 1, 2, 3, 7 and 8 were found in the litigating head injury group. Slick, Hopp, Strauss, and Spellacy (1996) found that scores on the F, F - K, and FBS were significantly correlated with performance on the Victoria Symptom Validity Test (VSVT), a test specifically designed to detect malingering; however, the FBS was the most strongly associated with neuropsychological test performance.

Schretlen and Arkowitz (1990) examined the strategies employed by prison inmates to fake either mental retardation or mental disorders. The performance of the inmates was compared to clinical groups of mentally retarded persons and psychiatric inpatients, and discriminant function analysis was used to predict group membership. Schretlen and Arkowitz found that both types of fakers (mental retardation and psychiatric disorders) demonstrated significantly higher F and F - K values than did non-faking groups. Furthermore, the scores of both faking groups on neuropsychological tests of malingering were also significantly worse than non-faking groups. Their findings supported the hypothesis that the use of a battery of test to detect malingering increased diagnostic accuracy over the use of a single measure, as they were able to correctly classify 95% of participants using discriminant function analysis of multiple variables.

The influence of litigation upon MMPI-2 profiles of minor head injury patients and moderate/severe head injury patients was investigated by Youngjohn, Davis and Wolf (1997). Youngjohn et al., found that in a group of 30 moderate or severe head injured patients referred to their clinic 18 were in ongoing litigation; however, in a group of 30 mild head injured patients all were in ongoing litigation. The investigators found that the litigating severe head injury group produced significant elevations on the MMPI-2 clinical scales 1, 3, and 8 relative to the non-litigating severe head injury group. Mild head injured patients produced elevated scores on clinical scales 1, 2, 3, and 7 relative to both the litigating and non-litigating severe head injury group. Youngjohn et al., noted the apparent paradoxical effect that litigation had on the profiles of less severely injured patients; that is, patients with less severe injuries produced the most significantly elevated profiles on the MMPI-2. The authors suggested that although litigation may have some effect on mild head injury patients, causing them to endorse more pathology, they hypothesized those patients with significant emotional difficulties or psychopathology may be more likely to pursue financial compensation.

Interestingly, they did not find any difference on the standard validity scales of L, F, or K between the different groups. However, their lack of findings may be due in part to the inconsistent utility of the F scale in detecting exaggerated profiles as well as their relatively small sample (N=30). The incorporation of other validity scales, particularly the FBS may have been appropriate for their participants, as it was designed to detect malingering in personal injury claimants.

Other investigators have also found cognitive and psychological malingering to produce independent response patterns on the MMPI-2 and neuropsychological tests of malingering (Greiffenstein, Gola, & Baker, 1995). Grieffenstein et al., divided participants who were referred for neuropsychological assessment into three groups, a traumatic brain injury group, a probable malingering group, and a persistent post-concussion group. The probable malingerers and post-concussion group were distinguished by the post-concussive group having returned to previous employment and not being involved in third-party lawsuits. That is, the persistent post-concussion group was involved in lawsuits to recover medical expenses, while the probable malingerers were seeking jury awards through third-party lawsuits. Using a known-groups method, Greiffenstein et al., found that domain specific measures were generally more sensitive to noncompliance than were MMPI-2 measures. The researchers found that the validity scales F, F - K, and O - S were more closely associated with a psychiatric malingering factor than a neuropsychological malingering factor. However, they did not incorporate the FBS in their analyses, which has shown some utility for detecting exaggeration of deficits in personal injury litigants.

The influence of litigation upon standard cognitive and personality measures has been a relatively consistent finding (Binder & Rohling, 1996). In a meta-analytic review of studies done on personal injury litigants, Binder and Rohling found a moderate effect size of 0.47 for mild head trauma. Of particular interest was the confirmation of previous findings of the paradoxical effect of litigation that abnormality and disability increased with less severe injuries.

Clinical Interview in Detecting Deception; Verbal and Nonverbal Indicators of Deception

The clinical interview is another standard feature of neuropsychological assessment, during which the clinician gathers information regarding the patient's complaints, premorbid functioning, and details of the incident in question. The clinical interview provides an opportunity for the neuropsychologist to compare the patient's presenting complaints with known findings from neuropathological examinations (i.e., electroencephalogram, computed tomography scan) to evaluate the presence of inconsistencies (Franzen, Iverson, & McCracken, 1990). Although some authors have questioned the utility of the interview for detecting malingering of psychiatric disorders empirical evaluation of the usefulness of the interview in detecting neuropsychological malingering has not been undertaken (Ziskin, 1984; Franzen, Iverson, & McCracken).

Research in social psychology laboratories has investigated nonverbal behaviors that are associated with lying; however, the application of this data has yet to be systematically applied to the detection of malingering in clinical psychology contexts. Although structured interviews have shown some promise for their utility in detecting different malingering strategies, the ability of clinical psychologists to be trained in detecting nonverbal indicators of lying and detecting false or exaggerated statements in a clinical interview has yet to be examined (Rogers, Gillis, Bagby and Monteiro, 1991).

Deception occurs when an individual attempts to convince another person to believe something that the deceiver considers false (Zuckerman & Driver, 1985). Folk wisdom purports that nonverbal means of communicating are of fundamental importance in detecting deception primarily because it is assumed that nonverbal behaviors are largely involuntary and hence more difficult to fake than are verbal messages. However, the deceiver may more easily control some nonverbal channels of communication than others. Ekman and Friesen (1974) found that relative to the face, the body tends to be a better source of cues that deception is occurring. Furthermore, an individual's tone and frequency of voice, provides additional cues that a person's communication is deceptive (Zuckerman, Amidon, Bishop & Pomerantz, 1982). The literature in the area of deception detection suggests that rather than a dichotomy of verbal versus nonverbal indications of deception, a hierarchy of communication channels with differing degrees of controllability exist for examining the veracity of a speaker (Zuckerman & Driver, 1985).

Reviews of the literature regarding the detection of deception suggest that naïve judges of deceivers achieve accuracy rates ranging from .45 to .60 (Zuckerman, DePaulo, & Rosenthal, 1981). Detection accuracy rates were improved when either bodily cues or verbal cues, such as tone and frequency of voice, were used rather than facial expressions, lending support to a hierarchy of communication channels model (Zuckerman et al., 1981). In explaining the behavioral cues associated with deceptive messages, two models emerged that appear to have that most relevance for the application of this area of research to detection of malingering on neuropsychological tests.

Physiological arousal has received some support as a mechanism through which behaviors indicating deception are produced (Zuckerman & Driver, 1985). Autonomic responses such as skin resistance and skin conductance differ when an individual is telling the truth or is being deceitful (Lykken, 1974; Waid & Orne, 1981). Furthermore, certain behaviors are considered to be associated with deception specific arousal, such as pupil dilations, voice frequency, eyeblinks, speech errors and speech hesitations (Zuckerman & Driver, 1985; DeTurck & Miller, 1985).

Cognitive factors also have a potential role in the production of behaviors that indicate deception. When an individual is required to lie on an experimental task, he or she must monitor several channels of communication while trying to convince a judge that their message is truthful. Depending upon the type of deception required of the participant, this task can become quite complex and cognitively taxing. Such an increase in cognitive tasks may lead to the production of such behavioral indicators of deception as speech errors and pauses, longer response latency, and pupil dilation (Zuckerman & Driver, 1985). Some behavioral indices, such as pupil dilation may be accounted for by more than one theoretical model and therefore the underlying mechanism for their production during deceptive communication may not be discernable (Zuckerman & Driver, 1985).

A meta-analytic review of the literature on correlates of deception found that eight verbal and nonverbal behaviors distinguished reliably between lie telling and truth telling (Zuckerman & Driver, 1981). Of primary relevance to the current study were the following identified behaviors: adaptors (behaviors unrelated to verbal content; e.g., scratching), speech hesitations and voice frequency. The effect sizes for these indicators ranged from *d*=.38 for adaptors, *d*=.54 for speech hesitations, and *d*=.68 for voice frequency which represent small to medium effect sizes (Zuckerman & Driver, 1981). Although, negative statements (i.e., frequent disparaging statements) and irrelevant information were identified as indicators of deception, the analysis of the specific content of deceptive communication made by neuropsychological malingerers is beyond the scope of the current investigation. DeTurck and Miller (1985) identified hand gestures, leg/feet gestures, and response latency as additional indicators that significantly contributed to the identification of deceptive communication.

DeTurck and Miller (1985) found that persons instructed to deceive demonstrated significantly greater sympathetic arousal than did non-deceivers. Furthermore, they found that adaptors, hand gestures, speech errors, and response latencies increased in frequency and intensity among deceivers relative to non-deceivers. DeTurck and Miller also found that aroused deceivers could be distinguished from aroused truth tellers on the basis of the six verbal and nonverbal behaviors they identified, suggesting that these behaviors are specific to arousal induced by deception rather than being unique to general arousal.

One concern regarding the application of nonverbal and verbal behaviors to the detection of malingering is the process of training observers to reliably identify these behaviors and make judgments. In a recent study by Kassin and Fong (1999) college students were given training in verbal and nonverbal behaviors for identifying deceptive communications; however, the accuracy of their judgments were no better than untrained controls, despite their having more confidence in their judgments. However, DeTurck, Harszlak, Bodhorn and Texter (1990) found that training in nonverbal indicators of deception significantly enhanced the accuracy of social perceivers in detecting lies. Furthermore, psychologists specifically interested in deception were more accurate in identifying lying than were other groups of psychologists (Ekman, O'Sullivan, & Frank, 1999). One factor that contributed to the different findings was differences in the training. Kassin and Fong provided portions of police interrogation training, while the other studies provided training focused on identifying specific behaviors.

Accurate judgment of communication as being either truthful or deceptive is also affected by familiarity with the deceiver (Brandt, Miller, & Hocking, 1980). DeTurck and Miller (1985) suggest that in order for a judge to perceive another person's verbal and nonverbal behaviors and make an accurate decision about the presence of deception, exposure to the deceiver's unaroused truthful communication is necessary to provide a baseline against which subsequent behavior is evaluated. The lack of exposure to a previous baseline may account for some of the findings of poor judgment accuracy. Although a baseline of unaroused and truthful communication may be more difficult to obtain in clinical practice of evaluating persons who may be malingering, it seems possible that an initial period of interaction prior to the beginning of the evaluation may be slightly less arousing than the actual examination process. In other words, until a patient is asked specific questions regarding their disability, less deception specific arousal should be present.

The motivation of participants to lie successfully is another factor that has an impact on how easily they are detected. Many studies in this area have utilized monetary awards for successful performance in being deceptive (Zuckerman, DePaulo, & Rosenthal, 1981). In their meta-analytic examination of research on lying with varying degrees of motivation, Zuckerman and Driver (1985) found that most visual behaviors (i.e., adaptors and blinking) showed a decrease in frequency and/or intensity when participants were highly motivated (offered a monetary reward) versus low motivated participants (no financial incentive). Furthermore, high motivation was associated with shorter response lengths, slower rates of speech, and increases in voice frequency than was low motivation. In other words, nonverbal behaviors were differentially impacted by changes in motivation to deceive.

If conditions of high motivation result in greater arousal than low motivation conditions, deceivers might be expected to engage in more eye blinks, more adaptors, demonstrate more speech hesitations, and increases in voice frequency. However, Zuckerman and Driver (1985) found that only voice frequency and eye blinks demonstrated the expected pattern. Zuckerman and Driver suggested that when deceivers are highly motivated they might expend more effort at controlling their behavior, which results in a reduction of some indicators of deception, such as body movements and length of communication. Another possibility is that simply increasing the reward for successful deception does not necessarily increase arousal during the task; rather, arousal may also be affected by possible punishments if caught.

DeTurck and Miller (1985) suggested that high motivation, in the form of monetary awards, might produce enough justification to reduce dissonance arousal and thus diminish behaviors associated with deception specific arousal. However, in the examination of malingerers, high motivation in the form of jury awards or reduced responsibility is of particular interest. Furthermore, DePaulo, Lanier and Davis (1983) found that when motivation was manipulated, highly motivated participants were more accurately judged from their nonverbal behavior, whereas low motivation participants were more accurately judged from their verbal behavior. Given the difficulty in approximating the high motivation of actual malingerers to succeed at deceit for a rewarding payoff and to avoid the costly consequences of being caught, analysis of both verbal and nonverbal aspects of behavior are of particular importance for simulated malingering studies.

It is crucial to the investigation of verbal and nonverbal behaviors of deception to evaluate several channels of communication, as deceiver's ability to control certain aspects of behavior is variable (Ekman, O'Sullivan, Friesen, Scherer, 1991). Ekman et al., found that when several behaviors were combined (voice frequency and smiles) the hit rate for detecting deception was 86.4%, suggesting that while some deceivers may be able to control some nonverbal behaviors when lying, other indicators are apparent.

When considering malingering in a forensic neuropsychological setting, it is important to remember that the details of a supposed disability communicated during an interview by a person who is malingering, probably does not represent a completely false statement. That is, an individual who has sustained a mild head injury is likely aware of the symptoms they have experienced and in reporting them to the evaluating neuropsychologist may exaggerate them without manufacturing untrue symptoms; thus, it may be particularly difficult to detect malingering through behavior analysis. However, malingering requires both deliberateness and the pursuit of an external reward to be diagnosed (Sweet, 1999). It is the deliberate attempt to deceive in addition to the possibility of receiving a reward that should contribute to the increase in arousal felt by persons intending to malinger, which should result in an increase of the behaviors that have been shown to be effective in detecting deception.

Aims of the Current Study

The purpose of the current study is to investigate the application of multiple techniques to the detection of deception on neuropsychological assessment. Of particular interest is whether or not evaluation of verbal and nonverbal indices of deception are detectable during a simulated clinical interview with participants instructed to believably fake the neuropsychological sequelae of head injuries.

The first stage of this study involves the application of a magnitude of error strategy to detect malingering. The magnitude of error strategy is based on the theory that simulators respond in a characteristically different manner than do impaired patients or normal controls (Rogers, Harrel, & Liff, 1993). Specifically, the magnitude of error strategy predicts that since malingerers are aware of the correct answers on some neuropsychological tests, they intentionally give inaccurate responses in order to appear more impaired. The analysis of the qualitative differences in patterns of responding among participants instructed to fake versus those that are instructed to do their best should reveal that fakers are more likely to provide incorrect answers that are grossly incorrect. Although, some research has found malingerers to provide more "near miss" errors, approximate answers are also observed in honest responders suggesting that grossly incorrect responses may improve discrimination more than "near misses" (Rogers, et al.).

The application of this technique is best suited to neuropsychological tests that have forced-choice answers such as the recognition trial of the California Verbal Learning Test (CVLT) and Raven's Progressive Matrices (RPM). The first step in this approach involves the determination of probabilities of selection for the items contained in the measure. Martin, Franzen and Orey (1998) accomplished this on an adapted recognition trial for the Wechsler Memory Scale- Revised, Logical Memory subtest by having graduate students, and faculty rate the probability of being selected for each item on a multiple-choice recognition trial. After averaging across each rater they arrived at probability values and were able to achieve inter-rater reliability of .85.

One advantage of the magnitude of error approach is that it has the potential for generating possible "critical items" which may be useful in identifying possible malingerers in a clinical setting. Furthermore, it incorporates already used measures and does not add significantly to the length of the neuropsychologcial test battery. It is hypothesized that participants instructed to malinger will select more low-probability items than will controls on the CVLT and Raven's Progressive Matrices, and the use of the magnitude of error approach will offer additional classification accuracy beyond that found with a calculation of recognition hits alone.

The application of research from social psychology regarding deceptive communication is also of interest in evaluating the utility of the magnitude of error approach. Specifically, it has been hypothesized that the process of intentionally deceiving another can be cognitively challenging (Zuckerman & Driver, 1985). In the current study this seems particularly relevant, as participants will be asked to convincingly simulate a brain injury, and be given specific information about the symptoms associated with this condition. Zuckerman and Driver hypothesized that the cognitively taxing nature of deception can lead to the production of such behavioral indicators as longer response latency.

When participants are attempting to simulate deficits, they must monitor their performance to ensure that they are appearing impaired, while still seeming believable, which should be a cognitively complex task. The magnitude of error strategy would predict that fakers would be more likely to deliberately select grossly incorrect responses. The additional decision making process of the individual attempting to fake should lead to increased response latencies due to the additional cognitive resources used to monitor their performance. Therefore, it is hypothesized that by recording responses of participants, significant increases in response latency will be demonstrated by the malingering groups, relative to controls particularly on forced-choice test items.

It is also one of the aims of this study for investigating the utility of the MMPI-2 for differentiating between controls, and individuals given instructions to fake specific cognitive dysfunction. Some investigators have found that faking on the MMPI-2 and on neuropsychological tests represent different kinds of attempts to deceive (Greiffenstein, Gola, & Baker, 1995).

However, certain scales such as the FBS have not been routinely included in the analysis of the MMPI-2, which may have contributed to the discrepancies regarding the MMPI-2's utility in detecting neuropsychological malingering in the literature. The FBS has shown some promise for use as a means of detecting exaggerated performances particularly with personal injury litigants (Lees-Haley, English, & Glenn, 1991; Larrabee, 1998). Furthermore, other researchers have found that participants instructed to fake mental retardation also demonstrated severe emotional disturbance (Schretlen, & Arkowitz, 1990) suggesting that some malingerers do not differentiate between cognitive and emotional impairment well.

In the current study, performance of the two groups on the MMPI-2 is of particular interest. The Faking Bad Scale (FBS) is hypothesized to add significantly to the classification accuracy of participants given instructions to fake a head injury. The items selected for this scale are primarily associated with somatic complaints that are endorsed by patients with minor head injuries, suspected to be exaggerating impairment. The FBS also consists of items that provide information about the vehemence with which participants attempt to portray themselves as being honest.

It is also of interest for the current study to determine the extent to which verbal and nonverbal information that is conveyed in a mock clinical interview is useful in classifying individuals who are attempting to simulate a head injury. In clinical situations individuals seeking evaluations of head injuries have potential rewards to gain. In the case of head injuries, the reward is often in the form of excuse from work or financial awards by juries.

The clinical interview is a standard part of most neuropsychological evaluations and may provide the trained observer additional cues to the possibility of symptom exaggeration, through the observation of behaviors associated with deception. The application of the analysis and observation of behaviors associated with deception to actual clinical populations is beyond the scope of the current investigation. Furthermore, it is important to realize that malingering in clinical populations exists on a continuum of persons deliberately trying to fake an injury to receive monetary awards, to individuals who are less aware of their reduced motivation to perform at the best of their ability on neuropsychological evaluations. Individuals, who are less aware of their reduced motivation to perform poorly, may experience less arousal while discussing symptoms with a neuropsychologist; hence, these individuals may not display the behavioral signs associated with deception. However, just as with any currently used measure of malingering, just because an individual does not score below a cut-off score does not mean that he is not malingering, nor does the presence of a single questionable performance indicate that he is.

Observation and analysis of behaviors associated with deception during the clinical interview may be useful in detecting certain approaches to malingering, specifically persons intentionally attempting to fake may be detected in this way. It is the aim of this investigation to determine if ratings of nonverbal behaviors by trained observers are able to accurately classify individuals instructed to feign cognitive impairment and individuals instructed to perform to the best of their ability.

The analog malingerers in the current investigation will be offered a financial incentive for producing a successful deceptive performance in order to simulate the incentive that head injured individuals might receive from a jury. Some have questioned the effect offering financial incentives has on deceptive performances (DeTurck and Miller, 1985). It is the specific aim of this study to attempt to replicate real world situations in which malingering may occur, by offering a financial reward for successfully appearing impaired without being detected as faking and by trying to simulate an actual neuropsychologcial evaluation. Furthermore, the possibility of receiving a substantial financial incentive has been demonstrated to result in increased autonomic arousal, which should result in an increase in behaviors indicative of deception (Gustafson & Orne, 1963).

Detection of deception through observation of verbal and nonverbal behaviors can be improved by the provision of a baseline observation of an individual's behaviors while less aroused (Brandt, Miller, & Hocking, 1980). Although someone attempting to malinger on a neuropsychological evaluation is likely to experience autonomic arousal throughout the investigation, it is likely that the level of arousal will increase during specific questions asked during the interview in regard to symptoms and the individual's experience of their supposed cognitive deficits. In order to investigate this hypothesis, participants in the current investigation will be video and audio taped during a brief screening interview prior to their participation in a mock neuropsychological examination. It is hypothesized that participants instructed to malinger will demonstrate increased voice frequency, response latencies, adaptors, and foot/leg movements, during a mock clinical interview than during their initial screening. Additionally, it is hypothesized that the malingering group will produce significant increases in the aforementioned behaviors relative to controls.

The hypotheses are:

<u>Hypothesis 1</u>. Participants in the malingering group will select more low-probability items than will control participants on the CVLT and Raven's Progressive Matrices.

Previous research by Martin, Franzen and Orey (1998) has found that malingerers were more likely to select low probability items than controls on a forced-choice memory task. Other experimenters have suggested that the low-probability item method may be applied to other neuropsychological tests. Support for this hypothesis would demonstrate the utility of the low-probability approach on commonly used neuropsychological measure of memory and of nonverbal reasoning. Low-probability items have not been investigated on the CVLT or RPM, two commonly used neuropsychological test. Furthermore, low-probability items, if effective in detecting malingering, can lead to the generation of "critical items" that can quickly alert the examiner to the possibility that faking is occurring and additional scrutiny is needed.

Hypothesis 2. Participants in the malingering group will demonstrate longer response latencies on the recognition trial of the CVLT and on Raven's Progressive Matrices relative to control participants.

Research in the area of nonverbal indicators of deception has demonstrated that longer response latencies are typically seen during deceptive communication (Zuckerman & Driver,1981; DeTurck & Miller 1985). The additional cognitive load of selecting the correct answer, monitoring if enough items have been missed, and making a decision as to whether or not to give an incorrect answer and which incorrect answer to give should require additional time for a response in participants attempting to fake a brain injury. The present study is unique in that it involves applying analysis of nonverbal behaviors to evaluating malingering on neuropsychological tests.

Hypothesis 3. Participants instructed to fake a head injury will produce higher scores on the FBS of the MMPI-2 relative to normal controls.

The Fake Bad Scale (FBS) has been found to be a useful measure in detecting malingered performance on neuropsychological tests (Slick, Hopp, Strauss, & Spellacy, 1996). Incorporating the MMPI-2 into a test battery provides a more approximate testing situation and allows for an analysis of the self-report styles of malingerers versus normal controls. Furthermore, the use of self-report measures in conjunction with other measures of functioning should lead to an overall improvement in detection accuracy, especially when those self-report measures are designed specifically to detect a response style consistent with malingering.

<u>Hypothesis 4</u>. Participants instructed to malinger will demonstrate an increase in voice frequency, adaptors, and foot/leg movements during a clinical interview relative to their demonstration of these behaviors during a screening interview and demonstrate a greater increase in these behaviors than control participants.

The application of nonverbal indicators of deception found in social psychology literature to the clinical interview represents a novel approach to detecting malingering. Voice frequency, adaptors and foot movements have been found to have moderate effect sizes in detecting deception (Zuckerman & Driver,1981; DeTurck & Miller 1985). Additionally, the use of a baseline evaluation should add to the ability of these variables to discriminate malingered performances (Brandt, Miller, & Hocking, 1980).

<u>Hypothesis 5</u>. The incorporation of a multiple-method approach (e.g., test performance, self-report, nonverbal behaviors) to detecting malingering should lead to improved accuracy of classification than the use of any particular dependent variable. Furthermore, the number of low-probability items selected on the CVLT and Raven's Progressive Matrices will offer additional classification accuracy in a discriminant function analysis beyond that found by raw scores on CVLT recognition or total number correct on the Raven's Progressive Matrices.

The use of multiple measures and detection strategies has offered the greatest promise in improving detection rates of malingering (Rogers, Harrell, & Liff, 1993; Nies & Sweet, 1994; Iverson & Franzen, 1996). Malingering represents a cognitively complex task, making successful dissimulation across several avenues of analysis particularly difficult to achieve. The inclusion of an analysis of the magnitude of error strategy has not been applied to the CVLT or RPM; however, for this approach to have an additional utility over simply examining total scores, it should provide classification accuracy beyond the total scores on these measures.

The hypotheses of this experiment were analyzed through independent samples t-tests, within subjects ANOVA, and discriminant function analysis. The dependent variables that were analyzed through one-tailed t-tests for the two groups are number of low-probability items selected on the CVLT recognition and RPM, response latencies on the

CVLT recognition and RPM, and FBS scores on the MMPI-2. Changes in average voice frequency from the screening interview to the clinical interview, change in adaptors from the first to second interview, and change in foot/leg movements from the first to second interview were analyzed by separate 2X2 within subjects ANOVAs. A step-wise discriminant function analysis was conducted to determine the rate at which participants are classified using the best discriminating variables, into either the malingering group or the control group.

METHODS

Participants

100 undergraduate students were recruited to participate in the study through the human subjects pool at Michigan State University and received partial course credit for their participation. 10 participants did not complete the neuropsychological exam and were subsequently removed from the analysis. 6 participants were removed from the analysis for endorsing a history of significant head injury and/or participation in litigation regarding an accident injury. 30 participants were assigned to the

control group and 54 subjects were assigned to the experimental group. Participants were assigned on a variable assignment schedule. No other data were trimmed.

The average age of participants was 21.4 years (SD = 3.8). The mean education level was 14.2 years (SD = 0.5). The participants consisted of 79% Caucasians, 16% African American, 4% Hispanic and 1% mixed; 60% of participants were female and 40% were male. No significant differences were found between control and experimental groups on ethnicity, gender, age, or education variables.

Procedure

Informed consent was explained to participants, which consisted of informing them of the nature of the instruments to be used, the fact that they would be audio and videotaped during parts of the experiment, and risks associated with participating. Participants were not informed about the purpose of the audio/video taping procedure prior to participation. Participants were then assigned to one of two groups. Participants received a set of instructions that explained how they were expected to perform on the administered tests. Control group participants were instructed to do their best, answer honestly and told that this is an evaluation of tests of cognitive and emotional functioning.

Participants assigned to the head injury malingering group (Experimental Group) were given a scenario that describes them being in an automobile accident, where another driver was at fault. Included in this scenario was a brief informational statement about typical symptoms of mild head injuries and these participants were instructed to try and produce the most severe and believable deficits without making it obvious that they were faking. Participants in the experimental group were also informed that if they succeeded in appearing impaired but were not detected as faking through statistical analysis they would be awarded $5. The scenario also included additional information regarding the potential consequences if this was a real-world situation and they were detected as faking (e.g. not receiving the reward, being fined); however, participants were informed that no attempts to simulate actual punishments would occur.

After receiving their instruction set participants completed the MMPI-2. Participants were scheduled for the second phase of the experiment in which they completed a neuropsychological evaluation lasting approximately 90 minutes. Upon arrival for the neuropsychological exam portion of the experiment, informed consent was discussed again. The participants were then told that the experimenter would be asking them some questions to calibrate the computer. They were told that they should ignore the instructions they were given prior to the MMPI-2 while answering these questions and answer honestly. This was done in an attempt to establish a baseline of voice frequency, adaptors and foot/leg movements. After the baseline interview, the baseline interview the participants were given their instruction sets again and were told to read over them and given time to prepare for the neuropsychological testing.

After 10 minutes the examiner returned, asked if the participant had any questions regarding the instructions, and asked the participant to explain what they were being asked to do to ensure that the participants understood the given instructions. Another experimenter was then sent in to begin the neuropsychological exam, and participants were told that this experimenter was unaware of their group assignment, and the experimental participants were told to imagine that this examiner was evaluating them as part of their attempt to receive financial compensation for their injury. The second examiner then came in and administered the neuropsychological exam and clinical interview. The neuropsychological exam consisted of administration of the CVLT, RPM and a clinical interview.

The CVLT was administered first, and during the delay the RPM and clinical interview were conducted. The clinical interview consisted of questions regarding their injury and it's impact for experimental group participants and several benign questions about summer plans for control participants. Following the conclusion of the test battery, the participants were given a brief questionnaire to assess if they followed instructions and how they attempted to deceive the experimenter. They were then debriefed about the nature of the experiment. Participants were informed about the variables of interest to the experimenter and the reason for the audio/video taping procedure. Participants were informed that they no identifying information would be associated with their audio/video taped performance. Participants in the experimental group were informed that they would be paid for their performance if they were not detected as faking by statistical analysis. Participants were given course credit.

Measures

a. **California Verbal Learning Test** (CVLT; *Delis, Kramer, Kaplan, & Ober, 1987*)

The CVLT is a learning task that assesses verbal learning and memory. The CVLT measures the immediate, short delay, cued, long delay and recognition of presented information. The task consists of a list of 16 words from four semantic categories that is presented orally for five immediate-recall trials (List A). A second 16-item word list (List B) is presented once. Free and category-cued recall trials follow a free recall trial for List B. Free and cued recall trials as well as a recognition trial are presented following a 20-minute delay. Test-retest reliability for the CVLT was .59 with the normative sample, and further analyses of the normative sample have cited a coefficient alpha of .74 and split-half reliability as .63; criterion-related validity has been demonstrated by its correlation with the Wechsler Memory Scale of .66 (Delis, Freeland, Kramer & Kaplan, 1988).

The CVLT recognition trial is composed of items from six categories. Correct items or hits, List B items that share a category with List A items, List B items that are not categorically similar to List A, items from neither list but categorically related, neither list similar phonemically, and unrelated items from neither list. Three psychologists at a large mental health clinic having familiarity with the CVLT rated items from these categories as having low, medium or high probability of being selected by a normal adult taking the CVLT. Inter-rater reliability was assessed and correlation coefficient of .84 was achieved across the 44 recognition items. The unrelated items that did not appear on either list were consistently rated as low-probability items and each item was assigned a score of 1 in the calculation of the low-probability items score.

b. Raven's Progressive Matrices (RPM; Raven, 1960)

RPM is a multiple-choice paper and pencil test that consists of a series of visual matching and analogy problems. RPM was designed to assess abilities to form nonverbal constructs to solve complex problems. The test presents items with an increasing level of difficulty and requires the selection of the correct choice from 6 to 8 presented choices. Test-retest reliability of the RPM have been found to be .80, while studies of concurrent validity demonstrate a correlation of .70 with conventional tests of intelligence (Burke, 1985).

Additionally, the possible responses for each item on the first 3 subtests of the RPM were rated by psychologists familiar with the RPM as to the probability (either high medium or low) that each item would be selected in the administration of the test to a non-impaired individual. Inter-rater reliability was assessed and a correlation coefficient of .80 was achieved across the 36 items on sets A, B, and C. The responses that were on average rated as low probability items by the various raters were given a score 1 in the calculation of low-probability items score. The responses that fell on average in between low and medium probability of being selected were assigned a score of 0.5 in the calculation of the low-probability items score.

c. Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989)

The MMPI-2 contains 567 items that are either endorsed as "true", "false" or left blank. The MMPI-2 can also be administered using only 370 items, which contains all of the clinical scales and the L, F, and K validity scales. Participants were administered the shortened version of the MMPI-2 due to time constraints. The MMPI-2 was designed to aid in psychiatric diagnosis. This inventory is widely used in neuropsychological evaluations and contains several scales to assess the validity of a generated profile (Lezak, 1995). Test-retest reliability of the MMPI-2 scales has been reported as ranging from .58 to .92 and internal consistency values ranging from .34 to .87 according to normative sample data (Butcher et al., 1989). Of particular interest to this study was the participant's score on the Fake Bad Scale (FBS) (Lees-Haley, English, & Glenn, 1991). Convergent validity of the FBS has been established by a correlation of .40 with other measures of malingering (Slick et al., 1996). Only the FBS was scored during this experiment, and no clinical scales were calculated.

d. Voice Frequency

The complete utterances of participants were recorded and digitized during both the baseline interview and mock clinical interview using a microphone and laptop computer. The digitized signal was analyzed for fundamental frequency in Hz using Prevaricator V 1.0 software. Frequency values were averaged for the baseline and clinical interviews and the means were compared to determine change scores. Reliability analysis of voice frequency achieved a coefficient alpha of .78.

e. **Response Latency**

Responses of participants were recorded *during the* recognition trial of the CVLT and RPM using a *microphone* and laptop computer. Response latency was measured as the time elapsing between the end of the examiner's question and the beginning of the participant's response. Responses were recorded on a computer program and the precise end of *the examiner's* question and precise beginning of the participant's response were measured in milliseconds. Response latency reliability analysis revealed a coefficient alpha of .72 for the CVLT and .30 for the RPM. Split-half reliability for the RPM was .92.

f. **Adaptors, Foot/Leg Movements**

Videotapes (without sound) of participants were made during both the baseline and clinical interviews. Frequency of adaptors (e.g., scratching, or other grooming behaviors unrelated to what is being said) were coded by two independent judges blind to the experimental design and purpose. These judges also coded frequency of foot/leg movements. Decision rules for what constitutes an adaptor included any self-grooming or hand movements that involved another object (e.g., tapping the table, opening a water bottle). Any expressive hand gesture was not considered an adaptor. Adaptors that were continuous were counted as one movement until it stopped. If the same movement began again after a stop it was counted as an adaptor.

Decision rules for foot/leg movements involved any movement of the feet or legs that were observable. Continuous movement of the foot or leg were counted as one movement until the movement stopped. If the movement started again it was counted separately. Postural shifts that involved movement of the legs were also counted. Inter-rater reliability coefficients of .92 were achieved by the raters for both adaptors and foot/leg movements.

RESULTS

Table 1 presents mean scores, standard deviations and ranges corresponding to calculated low-probability scores (i.e., infrequently selected items on the CVLT and RPM), FBS scores from the MMPI-2, response latency on the CVLT and RPM, changes in adaptors (ΔAdaptors), foot/leg movements (ΔFoot/Leg), and voice frequency (ΔVoice Hz) from baseline to clinical interview.

		Controls	Experimental
Low-Probability	Mean (SD)	1.3 (1.17)	6.25 (4.69)
	Range	0 - 4.5	0 – 19
FBS	Mean (SD)	12.77 (4.27)	28.78 (7.41)
	Range	5 - 25	9 – 40
CVLT Response Latency	Mean (SD)	1.09 (.33)	2.25 (.99)
	Range	.70 - 2.21	1.1 - 6.5
RPM Response Latency	Mean (SD)	7.43 (3.15)	9.34 (6.0)
	Range	3.68 - 17.35	3.06 - 43.88
Δ Adaptors *	Mean (SD)	-2 (4.9)	4.57 (6.0)
	Range	-17.5 - 4.5	-8.5 – 20
Δ Foot/leg *	Mean (SD)	-1.9 (5.53)	4.84 (8.44)
	Range	-12.5 - 12.5	-12 – 36
Δ Voice Hz **	Mean (SD)	.79 (1.4)	.48 (1.51)
	Range	-2.35 - 5.38	-4.23 - 4.99

* Positive scores reflect an increase in observed behaviors from baseline to clinical interview
** Positive scores reflect an increase in voice frequency from baseline to clinical interview.

Table 1. Means, Standard Deviations, and Ranges For Experimental and Control Groups

Given the apparent differences between standard deviations between the control and experimental groups additional analyses were conducted using the *v* ratio, which is a comparison ratio of standard deviations. The v ratio is calculated by dividing the experimental group standard deviation by the standard deviation of the control group. In every case the experimental group demonstrated greater variability on the measures. Significance of the v ratio was determined by calculating confidence intervals with an error rate of 5%. The null hypothesis is that the standard deviations are equal in the population; hence the v ratio should be equal to 1. The difference in standard deviations between groups is significant if a v ratio of 1 is not contained in the confidence interval.

	Se	UB	LB	v	Sig.
Low-Probability	0.47	3.69	1.83	4.01	s
FBS	0.30	2.32	1.15	1.74	s
CVLT Response Latency	0.51	4.01	1.99	3.00	s
RPM Response Latency	0.33	2.54	1.26	1.90	s
Δ Adaptors	0.21	1.64	0.81	1.22	ns
Δ Foot/leg	0.26	2.04	1.01	1.53	s
Δ Voice Hz	0.18	1.44	0.72	1.08	ns

Table 2. *V* ratio, Standard Error, Upper and Lower Bound of Confidence Interval

Low-probability items, FBS scores, and response latencies all demonstrated significant differences in variance between the control and experimental groups at the

$p < .05$ level. The nonverbal variables of adaptors, and voice frequency were not significant, while foot/leg movements was just outside of the 95% confidence interval and represents a significant difference in variances between the control and experimental groups.

Hypothesis 1. Participants in the malingering group will select more low-probability items than will control participants on the CVLT and Raven's Progressive Matrices.

In order to evaluate differences between control and experimental groups on number of low-probability items selected, a one-tailed, independent samples t-test was conducted. Low-probability scores on the CVLT and RPM were summed and this score was used in the analyses. This hypothesis was supported. Low-probability items were found to be statistically significant between control and experimental groups, $t(82) = -5.68$, $p < .001$. As predicted, participants in the malingering group selected more low-probability items on the CVLT recognition trial and RPM sets A, B, and C than did participants in the control group. The strength of the relationship between group and selection of low-probability items as indexed by η was .53 indicating a strong effect size. The effect size as estimated by the *d* statistic was 1.45 for low-probability items.

Hypothesis 2. **Participants in the malingering group will demonstrate longer response latencies on the recognition trial of the CVLT and on Raven's Progressive Matrices relative to control participants.**

In order to evaluate differences between control and experimental groups on response latency on the CVLT Recognition trial and RPM trials A, B, and C, independent samples t-tests were conducted. This hypothesis was partially supported. Group differences were found to be statistically significant for response latency on the CVLT, t (82) = -6.20, $p < .001$. As predicted, participants in the malingering group took longer to respond to items on the CVLT Recognition trial. The strength of the relationship between group and CVLT response latency as indexed by η was .56 indicating a strong effect size. The effect size as estimated by the d statistic was 1.57 for CVLT response latency.

Contrary to predictions, the effect for experimental group did not achieve significance for RPM response latency, t (82) = -1.63, $p > .05$. On average participants in the malingering group demonstrated longer response latencies on the RPM, but this difference was not significant. The strength of the relationship between group and RPM response latency as indexed by η was .17 indicating a small effect size. The effect size as estimated by the *d* statistic was .40 for RPM response latency.

Further analyses were done with the RPM response latency using inference probability evaluation. Inference probability is a statistical technique that is used to determine the likelihood a Type II error has been committed when a directional hypothesis is made. Inference probability uses the point biseral correlation as the mean and calculates the area under the curve between a Z score of zero and the correlation found in the sample. This area is added to the area under the curve above the correlation to result in a probability that a Type II error was made in the decision to fail to reject the null hypothesis. This technique indicated a 95% likelihood that a Type II error was committed by failing to reject the null hypothesis. The true correlation of response latency on the RPM and group was positive, but was likely not captured due to insufficient power.

<u>Hypothesis 3</u>. Participants instructed to fake a head injury will produce higher scores on the FBS of the MMPI-2 relative to normal controls.

In order to evaluate differences between control and experimental groups on the FBS on the MMPI-2, a one-tailed, independent samples t-test was conducted. This hypothesis was supported. The overall effect for experimental group achieved significance for FBS, t (82) = -10.86, $p < .001$. As predicted, participants in the malingering group endorsed more items on the FBS than did participants in the control group. The strength of the relationship between group and FBS as indexed by η was .77 indicating a large effect size. The effect size as estimated by the d statistic was 2.65 for the FBS.

<u>Hypothesis 4</u>. Participants instructed to malinger will demonstrate an increase in voice frequency, adaptors, and foot/leg movements during a clinical interview relative to their demonstration of these behaviors during a screening interview and demonstrate a greater increase in these behaviors than control participants.

In order to evaluate differences between control and experimental groups on voice frequency, adaptors and foot and leg movements, 2x2 within subjects ANOVAs were conducted. Participants were not randomly assigned, which limits interpretations of results. Means and standard deviations for observations of behaviors for baseline and interview are reported below in Table 3.

		Controls	Experimental
Baseline Adapt	Mean (SD)	9.57 (5.55)	6.09 (4.08)
Baseline Foot	Mean (SD)	12.30 (8.94)	8.94 (4.84)
Baseline Freq	Mean (SD)	8.58 (1.53)	8.80 (1.64)
Clinical Adapt	Mean (SD)	7.57 (4.64)	10.67 (7.22)
Clinical Foot	Mean (SD)	10.40 (6.62)	13.71 (10.32)
Clinical Freq.	Mean (SD)	9.37 (1.13)	9.28 (1.18)

Table 3. Means and Standard Deviations For Experimental and Control Groups

This hypothesis was partially supported. A significant main effect was found for adaptors, which was not predicted, F (1,82) = 4.09, $p < .05$. Experimental and control groups demonstrated differences on observed adaptors during the baseline measurement, with participants in the malingering group demonstrating fewer adaptors. The strength of the main effect of adaptors as indexed by η was .19 indicating a small effect size.

A significant interaction was also found for adaptors (baseline/clinical) X group (malinger/control), F (1,82) = 26.15, $p < .001$. As predicted, participants in the malingering group demonstrated a greater increase in adaptors during a clinical interview than did control participants. The strength of the interaction as indexed by η was .48 indicating a large effect size. *ANOVA results* for adaptors are summarized in Table 4 and Figure 1 *below.*

Source	SS	df	Mean Square	F	Sig.
Adaptors	63.8922	1	63.89	4.01	0.05
Adaptors * Group	416.7493	1	416.75	26.15	0.00
Error (Adaptors)	1306.852	82	15.94		

Table 4. 2X2 Within-Subjects ANOVA for Adaptors

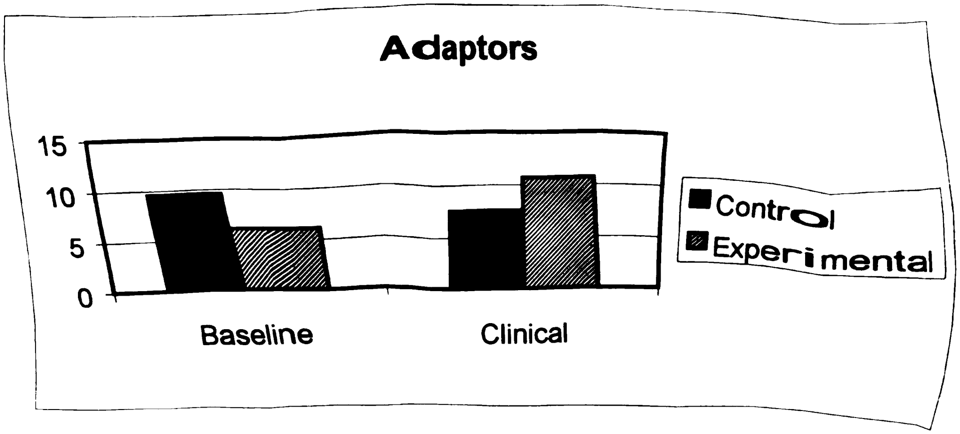

Figure 1. Adaptors Observed During Baseline and Clinical Interview

The main effect for foot/leg movements was not significant, F (1,82) = 2.76, p = .10. The strength of the main effect for foot/leg movements as indexed by η was .17 indicating a small effect size. A significant interaction of was found for foot/leg movements (baseline/clinical) X group (malinger/control), $F(1, 82) = 14.90$, $p < .001$. As predicted, participants in the malingering group demonstrated a greater increase in this behavior from their baseline relative to controls during a clinical interview. The strength of the interaction as indexed by η was .39 indicating a moderate effect size. ANOVA results for foot/leg movements are summarized in Table 5 and Figure 2 below.

Source	***SS***	**df**	**Mean Square**	**F**	**Sig.**
Foot/Leg	79.35	1.00	79.35	2.76	0.10
Foot/Leg * Group	428.81	1.00	428.81	14.90	0.00
Error (Foot/Leg)	2359.78	82.00	28.78		

Table 5. 2X2 Within-Subjects ANOVA for Foot/Leg Movements

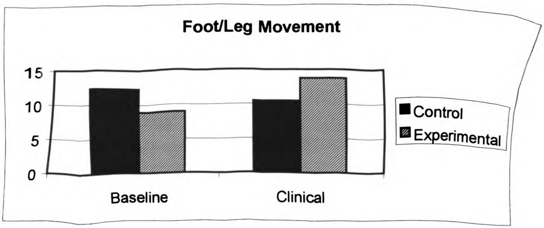

Figure 2. Foot/Leg Movements Observed During Baseline and Clinical Interview

A significant main effect was found for voice frequency, which was not predicted, *F*(1,82) = *14.12*, *p* < .001. Both experimental and control groups demonstrated significant increases in voice frequency during the clinical interview compared to their baseline levels. The strength of the main effect for voice frequency as indexed by η was .38 indicating a moderate effect size. Contrary to predictions, an interaction between frequency (baseline/clinical) X group (malinger/control) did not achieve significance, *F* (1,82) = .83, *p* = .36. On average participants in the malingering group demonstrated a smaller increase in frequency from baseline to their clinical interview than did controls, but this difference was not significant. The strength of the interaction as indexed by η was .09 indicating a small effect size. ANOVA results for voice frequency are summarized in Table 6 and Figure 3 below.

Source	SS	df	Mean Square	F	Sig.
Voice Frequency	15.57	1.00	15.57	14.12	0.00
Voice Frequency * Group	0.92	1.00	0.92	0.83	0.36
Error (Voice Frequency)	90.41	82.00	1.10		

Table 6. 2X2 Within-Subjects ANOVA for Voice Frequency

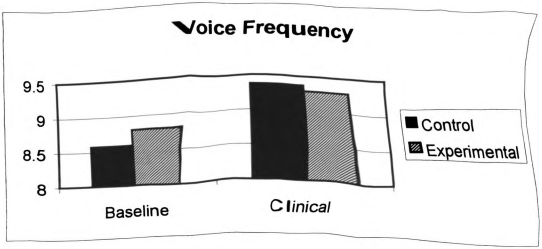

Figure 3. Voice Frequency During Baseline and Clinical Interview

Hypothesis 5. **The incorporation of a multiple-method approach (e.g., test performance, self-report, nonverbal behaviors) to detecting malingering should lead to improved accuracy of classification than the use of any particular dependent variable. Furthermore, the number of low-probability items selected on the CVLT and Raven's Progressive Matrices will offer additional classification accuracy in a discriminant function analysis beyond that found by raw scores on CVLT recognition or total number correct on the Raven's Progressive Matrices.**

This hypothesis was partially supported. A step-wise discriminant function analysis was conducted to determine the rate at which participants are classified into either the malingering group or the control group. The group variable degrees of freedom value for this two-group design is one; so one discriminant function was calculated. The single discriminant function was significantly associated with group membership [$X^2(4) = 97.99$, $p < .001$] and the canonical correlation coefficient related to this function was r = .84. Standardized discriminant function coefficients for the four variables included in the analyses following the step-wise procedure are presented in Table 7.

Standardized Canonical Discriminant Function Coefficients	
LOWPROB	0.340
FBS	0.734
CVLT Response Latency	0.420
Δ Adaptors	0.301

Table 7. Canonical Discriminant Function Coefficients

The classification accuracy of the discriminant function is shown in Table 8 below. The discriminant function achieved sensitivity, or a true positive rate for malingering of 92% and a specificity, or true negative rate for controls of 100%, with an overall classification accuracy of 95%. Furthermore, the use of the discriminant function results in a decrease in error rate of 86% compared to decisions of group membership based on sample size alone. These findings support the first part of this hypothesis in that use of several measures in a discriminant function analysis lead to increased accuracy in prediction of group membership.

	GROUP	Predicted Group Membership		Total
		control	experimental	
Count	control	30	0	30
	experimental	4	50	54
%	control	100	0	100
	experimental	7.41	92.59	100

Table 8. Discriminant Classification Results

In order to examine the hypothesis that low-probability items would add additional classification accuracy over raw scores on CVLT Recognition or number correct on the RPM, the raw scores were entered independently into the discriminant function and yielded a classification rate of 88%. When low-probability items were added to the discriminant function they did not contribute to the classification accuracy as predicted; thus the second part of this hypothesis was not supported. Analyses of correlations between variables revealed significant correlations of $r = -.91$ (RPM total) and $r = -.55$ (Recognition hits) with low-probability items suggesting that a significant proportion of the variance in the low-probability score is accounted for by these two variables. Point biseral correlations are included in column 1 of Table 9, and correlations between the variables are also reported in Table 9.

N=84	GROUP	Low PROB	FBS	CVLT RL	RPM RL	Δ Adapt	Δ Foot/leg	Δ Voice Hz	RPM Raw	RECOG HIT
GROUP	1									
LOW PROB	0.53**	1								
FBS	0.77**	0.43**	1							
CVLT RL	0.56**	0.41**	0.41**	1						
RPM RL	0.18	0.11	0.15	0.48**	1					
Δ Adapt	0.49**	0.17	0.46**	0.28*	0.32**	1				
Δ Foot/leg	0.40**	0.25*	0.37**	0.24*	-0.03	0.18	1			
Δ Voice Hz	-0.10	-0.15	-0.11	-0.27*	0.04	-0.1	-0.17	1		
RPM Raw	-0.59**	-0.91**	-0.49**	-0.38**	-0.15	-0.24*	-0.30**	0.14	1	
RECOG HIT	-0.60**	-0.55**	-0.52**	-0.50**	-0.22*	-0.23*	-0.34**	0.32**	0.54**	1
**	Correlation is significant at the 0.01 level (2-tailed).									
*	Correlation is significant at the 0.05 level (2-tailed).									

Table 9. Correlations

The intercorrelations between the variables of interest in this study revealed significant correlations ranging from .22 to .91. The relatively large correlations between variables likely contributed to the final discriminant function using only 4 variables (FBS, CVLT response latency, low-probability items, and change in adaptors) out of 7 entered in. Although the variables account for different methods of examining approaches to malingering, the overlap in variance accounted for limited the discriminating power of some of the variables during the step-wise procedure. When all of the variables are entered together, the contribution of each variable becomes more apparent. In Table 10 pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions are displayed. This matrix provides another way to study the usefulness of each variable in the discriminant function when all variables are entered simultaneously. The classification accuracy of the discriminant function remains the same.

Structure Matrix	Function
	1
FBS	0.727
CVLT Response Latency	0.415
Low-Probability	0.381
Δ Adaptors	0.343
Δ Foot/leg	0.263
RPM Response Latency	0.109
Δ Voice Hz	-0.061

Table 10. Pooled Within-Groups Correlations Between Variables and Canonical Values

DISCUSSION

The reliance of neuropsychology on self-report of symptoms and best effort on administered tasks demands that clinicians be vigilant to the influence of secondary gain on the data they collect during assessment of patients. Although multiple reviews of malingering research exist (Franzen, Iverson & McCracken, 1990; Rogers, Harrell & Liff, 1993; Nies & Sweet, 1994) no specific method for detecting dissimulation or exaggeration of symptoms has emerged as the most effective approach. The use of multiple methods to detect malingering provides the investigator with more evidence from which to make a more accurate decision about the likelihood that a patient is faking.

The purpose of the present study was to evaluate several methods for detecting deception and determine the utility of using a combined approach to improve specificity and sensitivity rates. Specifically, this study examined patterns of performance on neuropsychological tests, self-report of symptoms on the MMPI-2, and various nonverbal behaviors during a simulated clinical interview. The use of several methods for detecting deception resulted in high classification rates in the present study, with variables form each of the investigated methods contributing to classification accuracy.

The variability of performance by participants in the experimental group compared to the control group was significant. The differences between instructional set given to the experimental and control groups may account for part of this difference. The experimental participants were informed about several symptoms that may accompany head injury, but were also informed that most persons with brain injury do not exhibit all of these symptoms and were thus given considerable license to simulate neuropsychological deficits. Furthermore, similar variability has been demonstrated in other research with analog malingerers and suspected malingerers.

The increased variability of performance on neuropsychological and self-report measures by analog malingerers compared to normal controls is reflective of the differences in standard deviations seen in actual head injured patients attempting to exaggerate differences and controls (Ju & Varney, 2000; Berry et al., 1995). Interestingly, nonverbal indices of deception did not demonstrate the same discrepancy pattern between malingering participants and controls. Although foot/leg movements did demonstrate a significant difference in v ratio, the v ratio of 1.0 (null hypothesis) falls just outside of the upper bound of the confidence interval (1.01).

The pattern of variability on foot and leg movements was made clearer when they were examined at baseline versus during the clinical interview. For control participants, more variability of performance was seen during the baseline observation of foot/leg movements, while participants in the experimental group demonstrated less movement and less variability at baseline. This pattern was reversed during the clinical observation of foot/leg movements.

The variability seen in nonverbal behaviors may be due in part to methodological limitations of the current study. The baseline evaluation does not represent a baseline evaluation in a pure sense because it takes place after the participants have been assigned to groups. Although participants were told that the baseline questions were not part of the experiment and for calibration purposes, it is possible that the experimental group participants attempted to control their behaviors due to suspicion that this would affect future decisions about their veracity. Control participants may have been more anxious during the initial baseline evaluation, knowing that they were being recorded and been more relaxed by the passage of time prior to the clinical evaluation.

The relative lack of variability on certain nonverbal measures may be due to difficulties in controlling these indices by malingerers as they do their performances on neuropsychological and self-report measures. Previous research on nonverbal indices of deception has found a similar pattern of variability between interviewees telling the truth and those being deceptive (v = 1.32 adaptors; v = 1.12 foot/leg) (deTurck & Miller, 1985). Malingerers use a variety of approaches in their attempt to demonstrate their impairment, which underscores the need to utilize information from several channels in making determinations of the presence of malingering.

Hypothesis 1. Participants in the malingering group will select more low-probability items than will control participants on the CVLT and Raven's Progressive Matrices.

The low-probability items on the California Verbal Learning Test (CVLT) and Raven's Progressive Matrices (RPM) were more likely to be selected by participants given instructions to malinger than by control participants. However, low-probability items did not contribute to the discriminant function above and beyond the recognition hits score on the CVLT or total correct score on the RPM. In one of the initial studies of the method Martin, Franzen and Orey (1998) found that a magnitude of error approach on recognition portions of the Wechsler Memory Scale-Revised Visual Memory (VM) and Logical Memory (LM) contributed significantly to discrimination beyond recognition raw score values. Although the present study did find the use of low-probability items to have potential utility in detecting malingered performances, the clinical utility of this approach is called into question since it did not add to discriminating power beyond scores provided during routine administration (i.e., recognition hits and total score).

Further examination of the classification rate of low-probability items in the present study reveals that when low-probability score is used by itself, a 78% correct classification rate is achieved with a 10% false positive and 27% false negative rate. In comparison, when recognition hits are entered independently, an 81% correct classification rate is achieved with a 13% false positive and 22% false negative rate. Furthermore, the use of total score on the Raven's Progressive Matrices (RPM) results in only a 60% classification rate, yielding 40% false positive and 38% false negative rates. Essentially, the present study found that selection of low-probability items and raw scores on recognition measures perform similarly in discriminating between malingerers and controls, with recognition measures providing slightly better discrimination. Additional evidence for the similarity in these variables is found in the point biseral correlations. Correlations between low-probability items and the total scores were $r = -.91$ (RPM total) and $r = -.55$ (Recognition *hits*) respectively. Given the high correlation between the *variables it is* not surprising *that* no *significant increase* in *classification* accuracy was found, since they are essentially *accounting* for the same degree of variance.

The tests used in the present study differ from the nature of recognition tasks used in the Martin, Franzen and Orey (1998) study. Both LM and VM tasks used by Martin et al., presented participants with a several multiple-choice items including the correct answer as well as varying gradations of incorrect answers. In contrast, the California Verbal Learning Test (CVLT) recognition task presents correct items and varying gradations of incorrect answers, but each item is presented independently. The RPM does present a multiple-choice format, where the correct answer is presented within an array of incorrect answers, but this task is described as a problem-solving task rather than a recognition task. The differences between these tasks may account for the lack of additional discriminatory power of the low-probability score in the present study. Furthermore, the absence of a comparison group of actual head injured participants limits interpretations regarding the clinical utility of this approach with the CVLT and RPM.

The clinical usefulness of the magnitude of error approach is likely to be found in the generation of "critical items" rather than as a substitute for evaluation of raw scores. That is, items that are not selected by control groups or even head-injured participants and whose endorsement represents an increased likelihood of faking, may provide a quick and easy method to increase the examiner's awareness that an individual may be faking (Martin, Franzen & Orey, 1998). In the current study items defined as being low-probability items on the CVLT recognition trial were not selected by any of the control group participants, but 1 or more were selected by 32% of participants in the malingering group. Additionally, when items considered to be of low to medium probability were considered 63% of analog malingerers endorsed these items while none of the control group participants did. The use of critical items has also been supported in other research on neuropsychological malingering. DiCarlo, Gfeller, and Oliveri (2000) found that any errors on the first two subtests of the Halstead Reitan Neuropsychological Battery Category test were the most accurate indicator of malingering.

A comparison group of head-injured patients would be necessary in order to determine the utility of these "critical items" in differentiating malingerers from actual memory impaired individuals. The findings of the present study suggest that when certain items are endorsed on the recognition trial of the CVLT clinicians may be alerted to the increased possibility that an individual is malingering and can use additional screening measures to more fully evaluate an individual's motivation. Caution must be used however, in generating and reporting critical items. Use of critical items has received some support as a malingering detection approach (Tenhula & Sweet, 1996; Killgore & DellaPietra, 2000); however, reporting critical items in the literature represents an ethical dilemma, as these items could be readily obtained by individuals to aid in avoiding detection when being evaluated. Furthermore, actual patients may endorse critical items with certain neuropsychological conditions such as Wernicke-Korsakoff Syndrome, as they tend to confabulate on memory measures (Pachana, Boone, & Ganzell, 1998).

Hypothesis 2. **Participants in the malingering group will demonstrate longer response latencies on the recognition trial of the CVLT and on Raven's Progressive Matrices relative to control participants.**

Participants in the malingering group demonstrated greater response latencies on the California Verbal Learning Test (CVLT) Recognition trial than did controls as predicted. The additional cognitive steps required before responding was evident by the additional time taken by experimental participants. Although studies have examined response latency within the context of communicating deceptive messages, this approach has not been applied to detection of neuropsychological malingering.

Response latency on the recognition trials has some promise for improving detection of malingering, however, without a comparison group of actual head injured patients it is unclear as to the clinical utility of this approach. Slower response times would not be unexpected when examining individuals with symptoms that accompany legitimate brain injury (Lezak, 1985). Additional studies are needed to determine if malingering patients demonstrate longer response latencies than do legitimate head injured persons. Furthermore, a combination of the response latency and low-probability approach may yield *additional* discriminating information as to how malingerers *differ* from head injured patients.

Contrary to predictions, participants in the malingering group did not require additional time for responding on the Raven's Progressive Matrices (RPM) relative to controls. Inference probability analyses did suggest that the difference was in the predicted direction and the absence of a significant finding represents a Type II error. Furthermore, certain differences between the nature of the California Verbal Learning Test (CVLT) Recognition trial and the RPM likely contributed to the lack of findings.

The RPM is a forced choice test as is the CVLT recognition trial; however, the RPM is a novel task that requires a decision between 6 and 8 alternatives, whereas the CVLT Recognition trial simply requires a yes or no response. Individual differences of both the control and experimental participants in terms of scanning and abstract reasoning abilities likely contributed to the small differences between the groups. Furthermore, the purpose of the RPM is less readily apparent to the naïve participant, whereas the CVLT is an obvious memory task. It is possible that some participants in the malingering group were unsure of how to alter their performance on the RPM to appear as though they were suffering from a head injury.

The RPM may have some additional clinical utility in detecting malingering through an analysis of performance on the first third of items against performance on the last third. This "rate of decay" approach is based on the theory that malingerers have a difficult time accurately monitoring their performance and do not make more mistakes as the test items increase in difficulty (Gudjonsson & Shackleton, 1986). Although cross validation for this approach appears promising, it has not been validated on a clinical sample of probable malingerers (McKinzey, Podd, Krehbiel & Raven, 1999).

<u>Hypothesis 3</u>. Participants instructed to fake a head injury will produce higher scores on the FBS of the MMPI-2 relative to normal controls.

Participants in the malingering group demonstrated higher scores on the Faking Bad Scale (FBS) on the MMPI - 2 than did controls as predicted. The FBS was added to the battery in the current study in order to make the testing situation as close as possible to real-world conditions as well as to provide an additional avenue (i.e., self-report of symptoms) for detecting the possibility of malingering. The FBS consists of items that capture the particular response set of malingerers to report a variety of non-specific somatic symptoms while also presenting themselves as being honest and forthright (Lees-Haley, English, & Glenn, 1991). In the present study FBS represented the largest effect size and when the FBS was used independently in a discriminant function analysis it yielded a correct classification rate of 90%. It can not be determined from the present study how well the FBS performs with clinical populations of malingerers as opposed to analog malingerers; however, previous research has found the FBS to provide similar classification rates in clinical populations (Lees-Haley et al.; Larrabee, 1998; Martens, Donders & Millis, 2001).

<u>Hypothesis 4</u>. Participants instructed to malinger will demonstrate an increase in voice frequency, adaptors, and foot/leg movements during a clinical interview relative to their demonstration of these behaviors during a screening interview and demonstrate a greater increase in these behaviors than control participants.

The application of research from social psychology on deception seems to have potential for being applied to detecting malingering in clinical psychology. Behaviors such as adaptors and foot and leg movements were found to be sensitive to deceptive communication in the present study. As predicted, observed adaptors increased for participants in the malingering group but not for the control group. Additionally, foot and leg movements showed the same predicted pattern of increase from baseline to clinical interview for the experimental group but not for the control group.

One finding that was not predicted was group differences at baseline on adaptors and foot/leg movements. Control participants demonstrated more adaptors and foot/leg movements at baseline than did experimental participants. Both groups were expected to be approximately equal during the baseline evaluation, with the experimental group showing a relatively greater increase in behaviors. Methodological limitations may partially account for this finding. Participants were given instructions on how to approach the tasks prior to taking the MMPI-2, which was administered to a group. Upon their return for the individual neuropsychological evaluation, an attempt was made to record a baseline of nonverbal behaviors and participants were instructed to disregard their instructions while the instruments were calibrated. This design limits the interpretations that can be made in regard to the observed changes in nonverbal behavior.

However, considerable individual differences in nonverbal behaviors are not unexpected. The value of using a baseline approach to evaluating nonverbal indices of deception is that individual differences are accounted for and the reaction of the participant to the task of being deceitful can be more accurately measured. Future research would be able to address this methodological flaw by establishing a baseline of nonverbal behaviors prior to the assignment to groups.

Previous research has found that physiological arousal occurs during deceptive communication is manifested in certain verbal and nonverbal behaviors (deTurck & Miller, 1985). Both the experimental and control group in the current study were informed about being video and audio taped during the experiment, which likely created additional physiological arousal in both groups. However, the additional arousal created by having to come up with symptoms and details of an accident created the arousal specific to deception in the experimental group that is hypothesized to be responsible for the increase in adaptors.

Adaptors and foot/leg movements demonstrated potential for use in detecting malingerers using analog research design. When entered into a discriminant function independently, adaptors resulted in a 77% classification rate, while foot/leg movements resulted in a 71%

classification rate. The use of nonverbal behaviors to assist in detection of deception during a forensic evaluation may be of particular value because of the difficulty with which an individual is able to control these behaviors. However the utility of using nonverbal behaviors in clinical practice requires further research.

The present study attempted to improve detection rates by assessing a baseline of adaptors and foot/leg movements prior to a clinical interview based on research suggesting that familiarity with the deceiver's unaroused truthful communication should improve detection rates (Brandt, Miller, & Hocking, 1980). However, with clinical populations, physiological arousal is likely to be high throughout a forensic evaluation making it difficult to establish a baseline of these behaviors. Malingering in clinical practice exists on a continuum and it is likely that someone who is exaggerating symptoms of an actual head injury does not feel the same degree of physiological arousal associated with deception as someone who is generating symptoms of a head injury from scratch. Regardless, the use of observations of nonverbal behavior that have received empirical support may increase the reliability of clinicians' judgments of veracity during a clinical interview.

Contrary to predictions, voice frequency *did not* discriminate between malingerers and controls. *Previous* research with voice frequency has found effect *sizes of* *d*=.68 suggesting medium effect sizes in studies differentiating truth from deception (Zuckerman & Driver, 1981). Although the present *study* represents a novel application of previous social *psychology* research to *clinical* investigation, the *absence* of significant *findings is surprising.* One explanation for the lack of significant differences *between* control and experimental groups is the laboratory environment. Several extraneous sounds interfered with data collection, and recording equipment used was not sufficiently sophisticated to filter out background noise that may have impacted frequency measurements. No determinations regarding the utility of voice frequency in detecting malingering should be made from the present study. The factors that impacted the accurate recording of voice frequency should be addressed in future research and more sophisticated recording equipment would likely produce more accurate results.

From a clinician's point of view, voice frequency may be the least useable nonverbal indicator of deception during the clinical interview. While adaptors and foot/leg movements can be relatively easily observed, voice frequency requires additional equipment as well as an environment conducive to recording of voice. Additional research using a well-controlled environment adapting voice frequency to forensic clinical settings would provide more definitive information regarding the clinical utility of this approach.

Hypothesis 5. The incorporation of a multiple-method approach (e.g., test performance, self-report, nonverbal behaviors) to detecting malingering should lead to improved accuracy of classification than the use of any particular dependent variable. Furthermore, the number of low-probability items selected on the CVLT and Raven's Progressive Matrices will offer additional classification accuracy in a discriminant function analysis beyond that found by raw scores on CVLT recognition or total number correct on the Raven's Progressive Matrices.

The current research supports the use of multiple methods to improve detection rates of malingering. Most importantly in terms of clinical practice is the high sensitivity rate and specificity rates found in the present study. The importance of not mistakenly labeling an individual as having malingered when indeed they have given their best performance is paramount given the impact such a determination is likely to have. The application of the measures used in this study to clinical research involving actual head injured patients will provide the best data regarding how well the discriminant formula separates malingerers from head-injured patients. However, previous studies involving clinical samples have supported this approach (Rogers, Harrell, & Liff, 1993; Iverson & Franzen, 1996; Martens, Donders & Millis, 2001).

The complexity of malingering successfully requires an individual to monitor their performance across multiple domains to appear believable and avoid detection. By including self-report and nonverbal information observed during a clinical interview in conjunction with analysis of patterns of performance on neuropsychological instruments, malingering becomes increasingly difficult to achieve and easier to detect.

The performance of the four individuals who were malingering but were not detected by the discriminant function is of particular interest. Overall memory performance on the CVLT was below 1st percentile to the 8th percentile when compared to their same aged peers, yet their performance on the Recognition trial ranged from 12 to 15 out of a possible 16. All four participants reported approaching the task by being forgetful and acting as if they had difficulty paying attention. Two of the participants did not demonstrate any increase in nonverbal behaviors, while the other two did demonstrate an increase but not of sufficient size to be detected. Although the classification rates achieved using multiple methods were acceptable it is likely that despite the difficulty most participants had in maintaining their deceit across domains, it can be done. More importantly, no control participants were incorrectly identified as malingerers.

Despite fairly high rates of individual measures in differentiating between malingering and control participants in the present study, research with clinical populations has not supported the use of any particular measure as definitive in identifying exaggerated deficits. Use of the magnitude of error strategy on the CVLT and RPM yielded classification rate of 78% while the response latency of the CVLT resulted in an 85% classification rate. Studies with clinical populations have found similar classification rates using the CVLT performance patterns (Baker, Donders, & Thompson, 2000; Sweet et al, 2000).

The FBS produced an individual classification rate of 90% in the present study; however, recent research with actual head-injured patients found that the FBS was not able to sufficiently identify improbable response profiles independent of the CVLT and vice versa (Martens, Donders, & Millis, 2001). Low-probability items or FBS scores entered independently lead to higher false positive rates (13% and 7%, respectively). This finding reinforces the importance of using multiple measures and data sources to enhance decision accuracy in regard to malingering.

The incorporation of nonverbal behaviors observed during a clinical interview showed some promise in aiding in detection accuracy. Although the use of indices such as increased adaptors and foot/leg movements resulted in relatively high classification rates when used independently (77% and 71%, respectively), use of these indices alone resulted in unacceptable false positive error rates (56% and 63%, respectively). While use of observations during clinical interview may represent a rich source of data for determining effort, these observations lack sufficient specificity to be used independently.

Limitations of the Current Study

One of the primary limitations of the current study is the lack of a clinical comparison group. Although use of analog malingering groups is useful in determining the internal validity of detection measures, it does not address external validity. One of the paradoxes of this approach is that participants are asked to comply by faking, while actual malingerers fake their performance when asked to comply (Nies & Sweet, 1994). Furthermore, the use of positive incentives is easier to achieve than negative consequences using this paradigm due to ethical considerations. However, in real-world settings serious negative consequences often result from a determination of malingering ranging from discontinuation of medical benefits to fines or possible imprisonment (Rogers & Cruise, 1998). Rogers and Cruise found that participants who were given negative incentives produced more focused symptom presentations, suggesting that emphasis on possible negative outcomes may improve feigning performances in simulation designs. Furthermore, the $5 reward used in the present study was likely not sufficient motivation to improve performance. Larger rewards or alternatives such as raffle prizes may produce sufficient motivation without raising additional ethical concerns.

One solution to the limitations of the simulation design is analysis of convergence with findings from known-groups comparison designs. Although use of a known-groups comparison was beyond the scope of the current study, findings using established procedures (use of CVLT and FBS) are similar to those found in other known-groups research (Baker, Donders, & Thompson, 2000; Sweet et al, 2000 Lees-Haley et al.; Larrabee, 1998; Martens, Donders & Millis, 2001). Nonverbal indices of deception have yet to be examined in a known-groups design, and although they demonstrated some utility in the current study, their clinical usefulness has yet to be determined.

Despite the threats to external validity of the simulation design, college students may be more sophisticated in their attempts to malinger than actual head injured patients (Haines & Norris, 2001). Haines and Norris found that student malingerers were only identified at a rate of 26% when compared to actual head injured patients, whereas patients with history of head-injury asked to simulate symptoms were detected with 100% accuracy. However, Ju and Varney (2000) found that head injured patients could avoid detection of their attempts to malinger on the Portland Digit Recognition Test as well as non-clinical simulators. Although students appear to be a difficult group to detect as malingering, they differ from typical clinical populations on several demographic variables and further research using samples that more closely approximate likely referrals for forensic evaluation of head injury is needed.

Methodological limitations of the current study also warrant discussion. The participants were assigned to groups and given instructions about how to approach the task in the initial testing session. Participants then took the MMPI-2 and were scheduled for a follow-up appointment to complete the neuropsychological tests. When participants returned for the second session they were asked several questions to establish a baseline of nonverbal behaviors. Although they were told that this evaluation was not part of the experiment, the instructions they were previously given may have had an impact on their performance, thus limiting the conclusions that can be drawn from changes observed during the clinical interview.

Another limitation of the nonverbal data is the differences in time for the interview. Although efforts were made to equalize the time spent on the interview for the control and experimental groups, on average the experimental group interviews were slightly, although not significantly longer. This likely contributed to the increase in nonverbal behaviors seen. Participants in the control group were encouraged to elaborate their answers during the interview to equalize the amount of interview time between groups and this may also have had an effect.

Additionally, the lack of a true random assignment limits the interpretations from the current study. Although no differences were observed on demographic variables for the control or experimental group, the statistical analyses done in this study are based on the assumption of complete random assignment to groups.

Directions for Further Research

The current study suggests that incorporating objective techniques of measuring behaviors during clinical interviews has the potential for improving detection accuracy of malingering. A recent study by Vrij, Edward, Roberts and Bull (2000) evaluated two objective techniques for analyzing content of verbal communication and found that incorporation of these methods improved detection of deception over the use of nonverbal observations alone. Furthermore, future research should attempt to include clinical population samples to improve gerneralizability for findings.

Laboratory research can contribute to improving malingering detection; however, caution must be exercised in applying findings into clinical practice. Additional analog studies using more sophisticated and sensitive equipment for evaluation of voice frequency may provide more information regarding the utility of adapting this technique from social psychology research to the clinical realm. Also, laboratory settings that more closely control extraneous noise would eliminate potential confounding variables experienced in the present study.

Manipulations of the type of incentive given to malingerers also warrants additional research. Emphasis on negative consequences may alter the performance of persons trying to be deceptive, and attempts to more closely approximate the forensic setting would improve external validity of findings.

REFERENCES

Alfano, D., Neilson, P., Paniak, C., & Finlayson, M. (1992). The MMPI and closed-head injury. The Clinical Neuropsychologist, 6, 134-142.

Baker, R., Donders, J., & Thompson, E. (2000). Assessment of incomplete effort with the California Verbal Learning Test. Applied Neuropsychology, 7 (2), 111-114.

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

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

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

Bernard, L. C., McGrath, M. J., & Houston, W. (1993). Discriminating between simulated malingering and closed head injury on the Wechsler Memory Scale-Revised. Archives of Clinical Neuropsychology, 8, 539-551.

Berry, D. T. R., Wetter, M. W., Youngjohn, J. R., Gass, C. S., Lamb, D. G., Franzen, M. D., MacInnes, W. D., & Buchholz, D. (1995). Over-reporting of closed-head injury symptoms on the MMPI-2. Psychological Assessment, 7 (4), 517-523.

Binder, L. (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., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. American Journal of Psychiatry, 153, 7-10.

Binder, L. M., Villanueva, M. R., Howieson, D., & Moore, R. T. (1993). The Rey AVLT recognition memory task measures motivational impairment after mild head trauma. Archives of Clinical Neuropsychology, 8, 137-147.

Brandt, J., Rubinsky, E., & Lassen, G. (1985). Uncovering malingered amnesia. Annals of the New York Academy of Sciences, 444, 502-503.

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1980). The truth-deception attribution: Effects of familiarity on the ability of observers to detect deception. Human Communication Research, 6, 99-110.

Burke, H.R. (1985). Raven's Progressive Matrices: More on norms, reliability, and validity. Journal of Clinical Psychology, 41, 231-235.

Butcher, J. N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). Manual for administration and scoring the Minnesota Multiphasic Personality Inventory-2. Minneapolis: University of Minnesota Press.

Coleman, R. D., Rapport, L. J., Millis, S. R., Ricker, J. H., & Farchione, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. Journal of Clinical and Experimental Neuropsychology, 20 (2), 201-210.

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

Delis, D.C., Freeland, J., Kramer, J. H., & Kaplan, E. (1988). Integrating clinical asessment with cognitive neuroscience: Construct validity of the California Verbal Learning Test. Journal of Consulting and Clinical Psychology, 56, 123-130.

DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. Journal of Personality and Social Psychology, 45, 1096-1103.

DeTurck, M. A., Harszlak, J. J., Bodhorn, D. J., & Texter, L. A. (1990). The effects of training social perceivers to detect deception from behavioral cues. Communication Quarterly, 38 (2), 189-199.

DeTurck, M. A., & Miller, G. R. (1985). Deception and arousal: Isolating the behavioral correlates of deception. Human Communication Research, 12 (2), 181-201.

Diamond, R., Barth, J., & Zillmer, E. (1988). Emotional correlates of mild closed head injury: The role of the MMPI. International Journal of Clinical Neuropsychology, 10, 35-40.

DiCarlo, M.A., Gfeller, J.D., & Oliveri, M.V. (2000). Effects of coaching on detecting feigned cognitive impairment with the Category Test. Archives of Clinical Neuropsychology, 15 (5), 399-413.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. Psychological Science, 10 (3), 263-266.

Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer. (1991). Invited article: Face, voice, and body in detecting deceit. Journal of Nonverbal Behavior, 15 (2), 125-135.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. Journal of Personality and Social Psychology, 29, 288-298.

Faust, D. (1995). The detection of deception. Neurologic Clinics, 13 (2), 255-265.

Gass, C. (1991). MMPI-2 interpretation and closed head injury: A correction factor. Psychological Assessment, 3, 27-31.

Gass, C. & Russell, E. (1991). MMPI profiles of closed head trauma patients: Impact of neurological complaints. Journal of Clinical Psychology, 47, 253-260.

Gough, H. G. (1950). The F minus K dissimulation index for the MMPI. Journal of Consulting Psychology, 14, 408-413.

Gough, H. G., (1954). Some common misconceptions about neuroticism. Journal of Consulting Psychology, 18, 287-292.

Greiffenstein, M. F., Gola, T., Baker, W. J. (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. The Clinical Neuropsychologist, 9 (3), 230-240.

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

Guilmette, T. J., Hart, K. J., Giuiano, A. J, & Leininger, B.E. (1994). Detecting simulated memory impairment: Comparison of the Rey fifteen-item test and the Hiscock forced-choice procedure. The Clinical Neuropsychologist, 8 (3), 283-294.

Haines, M. E., & Norris, M. P. (2001). Comparing student and patient simulated malingerers' performance on standard neuropsychological measures to detect feigned cognitive deficits. The Clinical Neuropsychologist, 15 (2), 171-182.

Heaton, R. K., Smith, H. H., Lehman, R. A., Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. Journal of Consulting and Clinical Psychology, 46 (5), 892-900.

Iverson, G. L., & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. Journal of Clinical and Experimental Neuropsychology, 18 (1), 38-51.

Iverson, G. L., Franzen, M.D., & McCracken, L. M. (1994). Application of a forced-choice memory procedure designed to detect experimental malingering. Archives of Clinical Neuropsychology, 9 (5), 437-450.

Iverson, G. L., Franzen, M.D., & McCracken, L. M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. Law and Human Behavior, 15 (6), 667-678.

Ju, D., & Varney, N. R. (2000). Can head injury patients simulate malingering? Applied Neuropsychology, 7 (4), 201-207.

Kassin, S. M., & Fong, C. T. (1999). "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. Law and Human behavior, 23 (5), 499-516.

Killgore, W. D., DellaPietra, L. (2000). Using the WMS-III to detect malingering: Empirical validation of the Rarely Missed Index (RMI). Journal of Clinical and Experimental Neuropsychology, 22 (6), 761-771.

Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. The Clinical Neuropsychologist, 12 (2), 179-188.

Lees-Haley, P. R. (1991). MMPI-2 F and F - K scores of personal injury malingerers in vocational neuropsychological and emotional distress claims. American Journal of Forensic Psychology, 9 (3), 5-13.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI - 2 for personal injury claimants. Psychological Reports, 68, 203-210.

Lees-Haley, P. R. (1990). Contamination of neuropsychological testing by litigation. Forensic Reports, 3, 421-426.

Lees-Haley, P. R., & Fox, D. D. (1990). Neuropsychological false positives in litigation: Trail making test findings. Perceptual and Motor Skills, (70), 1379-1382.

Lezak, M.D. (1995). Neuropsychological assessment (3rd ed.). New York: Oxford University Press.

Lykken, D. T. (1974). Psychology and the lie detector industry. American Psychologist, 29, 725-739.

Martens, M., Donders, J., & Millis, S. R. (2001). Evaluation of invalid response sets after traumatic head injury. Journal of Forensic Neuropsychology, 2 (1), 1-18.

Martin, R. C., Franzen, M. D., & Orey, S. (1998). Magnitude of error as a strategy to detect feigned memory impairment. The Clinical Neuropsychologist, 12 (1), 84-91.

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

Millis, S. R., Putnam, S. H., & Adams, K. M. (1995, March). Neuropsychological malingering and the MMPI -2: Old and new indicators. Paper presented at the 30th Annual Symposium on Recent Developments in the use of the MMPI, MMPI- 2, and MMPI - A, St. Petersburg Beach, FL.

Millis, S. R., Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. Psychological Assessment, 7 (4), 463-471.

Nies, K. J. & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. Archives of Clinical Neuropsychology, 9, (6), 501-552.

Pachana, N. A., Boone, K. B., & Ganzell, S. L. (1998). False positive errors on selected tests of malingering. American Journal of Forensic Psychology, 16 (2) 17-25.

Raven, J. C. (1960). Guide to Using the Coloured Progressive Matrices. London: H. K. Lewis.

Rogers, R., Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. Law and Human Behavior, 22 (3), 273-285.

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

Rogers, R., Gillis, J. R., Bagby, R. M., & Montiero, E. (1991). Detection of malingering on the Structured Interview of Reported Symptoms (SIRS): A study of coached and uncoached simulators. Psychological Assessment, 3 (4), 673-677.

Schretlen, D., & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who fake insanity or mental retardation. Behavioral Sciences and the Law, 8, 75-84.

Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI - 2 Validity Scales. Journal of Clinical and Experimental Neuropsychology, 18 (6), 911-922.

Sweet, J. J., Wolfe, P., Sattlberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S., & Nies, K. J. (2000). Further investigation of traumatic brain injury versus insufficient effort with the California Verbal Learning Test. Archives of Clinical Neuropsychology, 15 (2), 105-113.

Sweet, J. J. (1999). Malingering: Differential diagnosis. In J. J. Sweet (Ed.). Forensic Neuropsychology: Fundamentals and Practice (pp. 255-285). Lisse: Sweets & Zeitlinger.

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., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. Journal of Clinical and Experimental Neuropsychology, 15 (4), 578-590.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. Journal of Nonverbal Behavior, 24 (4), 239-263.

Waid, W. M., & Orne, M. T. (1981). Cognitive, social and personality processes in the physiological detection of deception. In L. Berkowitz (Ed.). Advances in Experimental Social Psychology (Vol 14.). New York: Academic Press.

Wiener, D. N. (1948). Subtle and obvious keys for the MMPI. Journal of Consulting Psychology, 12, 43-47.

Wiggins, E. C., & Brandt, J. (1988). *The detection of simulated amnesia.* Law and Human Behavior, 12 (1), 57-78.

Youngjohn, J. R., Davis, D., & Wolf, I. (1997). Head injury and the MMPI-2: Paradoxical severity effects and the influence of litigation. Psychological Assessment, 9 (3), 177-184.

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

Zuckerman, M. & Driver, R. E. (1985). Telling lies: Verbal and nonverbal correlates of deception. In A. W. Siegman and S. Feldstein (Eds.), Multichannel Integrations of Nonverbal Behavior. New Jersey: Hillsdale.

Zuckerman, M., Amidon, M. D., Bishop, S. E., & Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. Journal of Personality and Social Psychology, 43, 347-357.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 14). New York: Academic Press.

APPENDIX

INSTRUCTION SET A

The experiment in which you are going to participate involves several commonly used neuropsychological tests and a personality inventory. We are trying to validate some new test items on the tests you will be administered. You are to perform to the best of your abilities on all of the neuropsychological tests you will be administered. Some of the items will be difficult while some may be easy for you. No one is expected to answer all items correctly. Just do your best. On the personality inventory, please answer the items honestly as they apply to you. Your response sheet will be identified by a subject number and your responses will be kept confidential. We are only interested in validating some test items, some please respond frankly. We appreciate your willingness to participate in our research project. One of the experimenters will be happy to answer any questions that you have at the conclusion of the experiment.

INSTRUCTION SET B

The experiment in which you are going to participate involves several commonly used neuropsychological tests, a personality inventory and a brief clinical interview. For the purposes of this experiment you are to imagine that you sustained a head injury in a car accident that was not your fault. Initially you had some memory and concentration problems, as well as headaches, which resolved a short time after the accident. Although these symptoms have resolved, imagine that you have decided to pursue a lawsuit against the other driver and hope to increase the size of your claim by faking brain damage. You are to try and fake common symptoms of head injury in the most realistic fashion in order to prevent detection by the examiner. Common symptoms of a mild head injury include:

- Memory problems
- Attention and concentration problems
- A general slowing of mental abilities
- Difficulty with problem solving
- Fatigue, sensitivity to light and noise, irritability
- Nausea, headaches, dizziness

If you are able to demonstrate impairment and avoid detection by the examiner, you will be awarded $5. However, if you are determined by the examiner to be faking, this could result in the loss of your court settlement ($5) and in a real-world situation would likely result in other consequences such as a severe court penalty.

In the first phase of this experiment you will complete a personality inventory. Please answer the questions as if you are pursuing litigation for a head injury and are trying to appear impaired by the injury, while trying to seem believable. In the second phase of this experiment you will be given a few neuropsychological tests of memory and problem-solving abilities and be interviewed by the examiner. During the interview you will be asked questions by the examiner regarding how the head injury has impacted your school or work, symptoms you have experienced, and your recollection of the accident. You will be better able to deceive the examiner if you have thought about your answers ahead of time.

We appreciate your willingness to participate in our research project. One of the experimenters will be happy to answer any questions that you have at the conclusion of the experiment.

PRE-SCREENING INTERVIEW

Note: This screening interview is *NOT part of the* experiment and this should be communicated to the subject. This should be explained as a few brief questions to calibrate the audio and video equipment to be used in the actual experiment.

1. How many experiments have you participated in so far?

2. What has your experience of the experiments been so far? What kinds of experiments have you participated in?

3. What is your major at MSU? How did you choose that major or what about it appeals to you?

4. Have you ever had a head injury before? If so, what happened?

5. Have you ever been involved in any kind of litigation (as the claimant or defendant)?

CLINICAL INTERVIEW
(Instruction Set A)

I am going to ask you some questions and record your answers

1. Why did you choose to participate in this particular experiment?

2. Why did you choose to attend MSU?

3. What has been your most interesting class and why?

4. What are your plans for the summer?

"CLINICAL INTERVIEW"
(Instruction Set B)

I have a few questions I would like to ask you regarding your accident.

1. Tell me what happened in the accident.

2. Please describe the symptoms you have experienced as a result of your injury.

3. In what ways have you noticed the effects of your injury in your school or work?

4. What would you like to find out as a result of this testing?