2004

Are we honestly studying malingering?: a profile and comparison of simulated suspected malingerers

Adrianne M. Brennan
Louisiana State University and Agricultural and Mechanical College

ARE WE HONESTLY STUDYING MALINGERING?
A PROFILE AND COMPARISON OF SIMULATED AND SUSPECTED MALINGERERS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

in

The Department of Psychology

by
Adrianne M. Brennan
B.A., University of New Orleans, 2001
December 2004

Table of Contents

List of Tables

List of Figures

Astract

Malingering research typically uses analog simulation designs or the differential prevalence design among "real" patients. Both have been criticized for methodological limitations in external and internal validity, respectively. Samples of simulated malingerers were compared to suspected malingerers to examine generalizability of analog findings. Overall results support the use of simulation designs. Furthermore, it was demonstrated that stringent selection of suspected malingerers maintains internal validity of the differential prevalence design. A second focus, to determine if demographic matching of simulated malingerers is necessary, showed that matching on age and race is not necessary.

Introduction

Malingering is a very costly issue in the United States accounting for nearly one-fifth of all medical care cases (Ford, 1983). The medical and legal costs of malingering are estimated to be over $5 billion annually (Gouvier, Lees-Haley, & Hammer, 2003). An estimated 18% - 64% of litigating neuropsychological patients are believed to be malingering (Binder, 1993; Heaton, Smith, Lehman, & Vogt, 1978).

Malingering research typically uses one of two possible designs, the analog simulation design or the differential prevalence design. The simulation design utilizes normal individuals who pretend as if they have brain damage. The differential prevalence design employs patients who are considered "at-risk" for malingering. Both designs have been criticized for methodological flaws (Shadish, Cook, & Campbell, 2002). The simulation design is criticized for external validity concerns, specifically for the unknown generalizability to forensic populations. The differential prevalence design is criticized for internal validity issues, particularly for the ambiguity in which "at-risk" malingerers are chosen. This study will attempt to address these concerns by directly comparing stringently-selected "at-risk" malingerers to simulated malingerers.

Definitions of Malingering

According to the DSM-IV-TR, the essential feature of malingering is "intentional or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs" (American Psychiatric Association, 2000).

Because the DSM-IV-TR does not consider malingering a specific disorder, guidelines rather than diagnostic criteria are provided. These guidelines include a (1) medicolegal context of presentation, (2) marked discrepancy between claimed disability and objective findings, (3) lack of cooperation in assessment and treatment, and (4) the presence of antisocial personality disorder (American Psychiatric Association, 2000). In response to the DSM-IV-TR's broad categorization, various criteria have been proposed to more precisely define malingering (Rogers, 1997; Trueblood & Schmidt, 1993; Greiffenstein, Baker, & Gola, 1994). Perhaps the most thoroughly outlined proposal is provided by Slick, Sherman, and Iverson (1999).

Slick et al. (1999) define malingering as "the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility." (p. 552). Furthermore, the authors describe three categories of malingering, namely *possible, probable*, and *definite*. For a patient to classify into one of these categories some combination of four criteria is to be met. The four criteria are:

- *Criterion A: Presence of a substantial external incentive* - at least one clearly identifiable and substantial external incentive is present at the time of examination.

- *Criterion B: Evidence from neuropsychological testing* - evidence of exaggeration or fabrication on neuropsychological tests as evidenced from at least one of the following:

    1.) Definite response bias - below chance performance (p<.05) on one or more forced-choice measures.

2.) <u>Probable response bias</u> - performance on a well-validated test or index is consistent with fabrication or exaggeration.

3.) <u>Discrepancy between test data and known patterns of brain functioning</u>.

4.) <u>Discrepancy between test data and observed behavior.</u>

5.) <u>Discrepancy between test data and reliable collateral reports.</u>

6.) <u>Discrepancy between test data and documented background history.</u>

- *Criterion C: Evidence from self-report* - significant inconsistencies or discrepancies in a patient's self-reported symptoms that suggest fabrication or exaggeration as evidenced by one of the following:

  1.) <u>Self-reported history is discrepant with documented history</u>.

  2.) <u>Self-reported symptoms are discrepant with known patterns of brain functioning</u>.

  3.) <u>Self-reported symptoms are discrepant with behavioral observations</u>.

  4.) <u>Self-reported symptoms are discrepant with information obtained from collateral informants</u>.

  5.) <u>Evidence of exaggerated or fabricated psychological dysfunction -</u> performance on well-validated validity scales or indices on self-report measures of psychological adjustment are strongly suggestive of exaggeration or fabrication.

- *Criterion D: Behaviors meeting necessary criteria from groups B or C are not fully accounted for by Psychiatric, Neurological, or Developmental Factors -*

behaviors are the product of an informed, rational, and volitional effort aimed at least in part toward acquiring or achieving external incentives.

To qualify as a *definite* malingerer, the patient must meet criteria A, B1, and D; meaning there must be substantial external incentive, the presence of a definite negative response bias on neuropsychological test(s), and no psychiatric, neurological, or developmental factor that would significantly diminish one's capacity to appreciate laws or mores against malingering.

To qualify as a *probable* malingerer, the patient must meet criterion A, two or more from B1-B6, and D, or criterion A, one from B1-B6, one from C1-C5, and D. Therefore, a patient can classify as a *probable* malingerer in two ways, by having the presence of external incentive, two pieces of evidence from neuropsychological testing, and no psychiatric, neurological, or developmental disorder, or by having external incentive, one piece of evidence from neuropsychological testing, one piece of evidence from self-report, and no psychiatric, neurological, or developmental disorder.

There are also two ways in which a patient can qualify as a *possible* malingerer, the patient must either meet criterion A, one from C1-C5, and D, (external incentive, evidence from self-report, and no psychiatric, neurological, or developmental disorder) or must meet criteria that would classify him/her as a *definite* or *probable* malingerer with the exception of criterion D. See Table 1.

While fairly new, the proposed definition and criteria of Slick et. al (1999) appear to be gaining support in the research community. Several recent studies have classified subjects according to this definition and criteria, demonstrating a strong conceptual framework from which

Table 1

Criteria and Classification of Malingered Neurocognitive Dysfunction of Slick, Sherman, & Iverson (1999).

Criterion A: Presence of a Substantial External Incentive
Criterion B: Evidence from Neuropsychological Testing
Criterion C: Evidence from Self-Report
*Criterion D Behaviors are not fully accounted for by Psychiatric, Neurological, or
        Developmental Factors*

Classification	Criterion A	Criterion B	Criterion C	Criterion D
Definite malingering	X	X*	(X)	X
Probable malingering	X	X (two pieces)		X
Or				
Probable Malingering	X	X (one piece)	X (one piece)	X
Possible Malingering	X		X	X

*Must Include Definite Negative Response Bias

to base a study (Mathias, Greve, Bianchini, Houston, & Crouch, 2002; Greve, Bianchini, Mathias, Houston, & Crouch, 2002).

<div align="center">Malingering Detection Methods</div>

Forced-Choice Measures

The most common malingering detection method is based on forced-choice (Lezak, 1983). A measure is presented in multiple-choice format and the patient's performance is compared to what would be expected by chance alone. (Rogers, 1997). For example, on a measure that consists of two items, the patient could theoretically answer 50% correctly just by guessing (Haines & Norris, 1995). The assumption behind this method is that if a subject scores significantly below chance ($p < .05$) there is purposive distortion (Reynolds, 1998). A major criticism of this method, however, is that of its low sensitivity. This method is extremely conservative and only the most blatant malingerers are caught. In response to this low sensitivity came the derivation of cut-off scores (Haines & Norris, 1995).

A cut-off score typically represents the lowest score achieved by subjects with documented brain damage. Therefore, if a patient with minor, or no, documented brain-injury performs significantly worse than the cut-off, malingering is to be suspected (Haines & Norris, 1995). Utilizing cut-off scores improves the sensitivity of the forced-choice method (Rogers, 1997), but at the cost of reduced confidence in the interpretation. Specificity is lowered because of the increase in false-positives.

One of the most widely used standardized forced-choice measures is the Test of Memory Malingering (TOMM) (Tombaugh, 2002). The TOMM is an objective, criterion-based measure

which discriminates between actual and feigned memory impairment. Its advantages include being insensitive to the effects of demographic variables, traumatic brain-injury, and neurological or psychological disease, its perceived difficulty exceeding its actual difficulty, and its high face validity (Tombaugh, 2002). Furthermore, it is psychometrically sound and has been shown to meet the Daubert Court standard of admissibility (Vallabhajosula, & van Gorp, 2001). Using a cut-off score of 45 appears to accurately identify the majority of simulating malingerers (Tombaugh, 1996).

Performance Curve

The performance curve demonstrates the higher frequency in which easy items are correctly answered compared to more difficult items (Rogers, 1997). In other words, it reflects the increasing proportion of committed errors when test-item difficulty is raised. This phenomenon can be used as a malingering detection technique. Evidence has shown that simulated malingerers do not generate the typical performance curve, that is they fail a "more-than-expected" proportion of easy items compared to their performance on more difficult items (Frederick & Foster, 1991).

A measure that relies on the performance curve is the Dot Counting Test (DCT). This measure presents stimuli of varying (and mixed up) difficulty levels to determine the consistency of an individual's response time and error-rate (Lezak, 1995). In non- malingering subjects, a positive correlation is expected between difficulty level and both time to respond and number of errors committed. Response time is assumed to increase with increased item difficulty, therefore more than one pronounced discrepancy raises the likelihood of exaggeration. In addition, error-

rate should be no greater than 2 items. A deviation from this pattern raises the suspicion of malingering. Overall, evidence supports error-rate as the stronger indicator of malingering (Frederick, 2002; Binks, Gouvier, & Waters, 1996).

Floor Effects

There are many problems and tasks that are easily accomplished by most individuals, including those with brain damage. Malingering detection utilizes this knowledge by examining floor effects. Floor effects are extremely low performances observed when malingerers misjudge the difficulty of easy tasks and perform more poorly than brain-damaged patients (Millis & Kler, 1995). A drawback to this method, however, is that it is sensitive to true memory impairment and correlates considerably with measures of cognitive competence (Vallabhajosula, & van Gorp, 2001; Lezak, 1995).

The Rey-Fifteen Item Memory Test (MFIT) is a commonly used measure that utilizes the floor effect (Frederick, 2002). This measure is sensitive to true memory impairment, therefore the cut-off score is not fixed. For comparison to non-clinical and psychiatric populations a cut-off score of 9 provides appropriate predictive accuracy; however if a differential diagnosis of amnesia or dementia is suspected, a cutoff score of 7 should be used (Goldberg & Miller, 1986; Bernard & Fowler, 1990; Frederick, Sarfarty, Johnston, & Powel,1994; Lezak, 1983; Lee, Loring, & Martin, 1992).

Validity Indices

Many self-report measures of psychological functioning contain validity scales meant to detect if respondents are answering in a manner which invalidates the overall results. More

specifically, these scales can indicate the direction of invalidation. For example, the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) has at least two indices which can be used for malingering detection. The F or "infrequency" scale measures the extent to which a person answers in an atypical and deviant manner. A score of 70 or above is suggestive of possible malingering. The Dissimulation or F-K index determines the likelihood and direction of exaggeration. A score of 12 or greater indicates a fake bad profile, while a score of -12 or less indicates a fake good profile (Groth-Marnat, 1997). Similarly, the Personality Assessment Inventory (PAI) contains scales appropriate for use in malingering detection. The strongest indicator is the Negative Impression Management scale which measures the degree to which an individual presents an exaggerated, unfavorable impression of distress. A score of 92 or greater is indicative of possible malingering (Morey, 2003).

<center>Issues in Malingering Research</center>

The vast majority of malingering research is based on the simulation design. This design utilizes non-clinical subjects, typically university undergraduates, asked to feign brain damage. This design is often criticized for external validity concerns, specifically its unknown generalizability to actual malingerers (Haines & Norris, 1995; Rogers, Bagby, & Dickens, 1992). One reason generalizability is questioned is because simulated malingerers do not have the same motivation that actual malingerers have to fake deficits. Without this motivation, it is possible that simulated malingerers may over-estimate the deficits associated with a mild head injury (Haines & Norris, 1995). One study attempted to remedy this issue by offering subjects who successfully faked deficits large financial incentives (Bernard, 1990). This, however, was not

successful and the simulated malingerers who received financial incentive differed only slightly from simulators with no incentive. Another concern is that of demographic matching. While most designs incorporate subjects who are demographically matched to known malingerers, there is no clear-cut evidence that supports its necessity.

The differential prevalence design is another paradigm often used in malingering research. This design utilizes patients considered "at-risk" for malingering. Participants are considered "at-risk" because they have a history of a mild closed head injury, no documented evidence of brain damage, and are actively involved in litigation (Tombaugh, 1996). Typically, inclusion criteria for "at-risk" malingering are very broad and do not ensure that any, or even one, subject is truly malingering. This poses a substantial threat to internal validity. If subjects are classified according to such broad criteria it is likely that non-malingerers will be included in the study, thus diluting the sample and reducing the effect size.

Currently, researchers must choose between using a "clean" homogeneous sample via the simulation design or a "dirty" heterogeneous sample via the differential prevalence design. Both designs may contain substantial threats to validity; however it may be possible to remedy these issues. Concerns with internal validity may be resolved by stringently selecting participants for the differential prevalence design. If participants were selected according to narrowly defined criteria, such as those provided by Slick et al. (1999), internal validity would be preserved. "At-risk" malingerers who met classification criteria of either *possible*, *probable*, or *definite* malingering could then be considered suspected malingers (and will be referred to as such for the remainder of the paper). Concerns with external validity can be answered by comparing

simulated malingerers to suspected malingerers, which would provide an estimate of

generalizability to forensic populations.

Purpose of Study

<u>Research Questions and Hypotheses</u>

<u>Question 1:</u> How accurately do simulated malingerers compare to true malingerers?

<u>Hypothesis 1a:</u> It is hypothesized that the frequencies of *possible* and *definite* classification will be significantly different for simulated and suspected malingerers.   Furthermore, it is predicted that the modal classification category of simulated malingerers will be *definite* whereas the modal classification of suspected malingerers will be *possible*. No difference on *probable* classification is expected. See Figure 1.

<u>Hypothesis 1b:</u> It is hypothesized that simulated malingerers will perform significantly worse on the TOMM compared to suspected malingerers.

<u>Hypothesis 1c:</u> It is hypothesized that simulated malingerers will perform significantly worse on the MFIT compared to suspected malingerers.

<u>Hypothesis 1d:</u> It is hypothesized that simulated malingerers will perform significantly worse on the DCT compared to suspected malingerers.

<u>Question 2:</u> Will predictions remain constant in cross validation?

<u>Hypothesis 2:</u> It is predicted that frequencies observed in the first study will not be significantly different in a second sample.

<u>Question 3:</u> Is demographic matching of simulated malingerers to true malingerers necessary?

<u>Hypothesis 3:</u> It is hypothesized that the frequencies of *possible*, *probable*, and

definite classification will not be significantly different for demographically

matched and unmatched simulators.

Figure 1

<u>Expected Modal Frequency of Classification Category in Known and Simulated Malingerers:</u>

	Simulated				Suspected		

1 = Definite
2 = Probable
3 = Possible
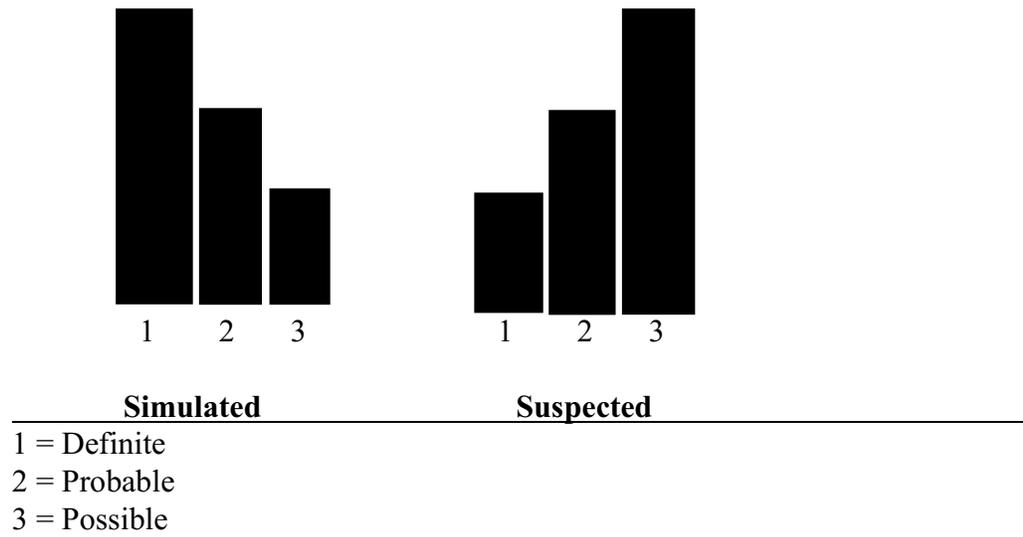
Method

Participants

A power analysis was performed to determine the number of subjects needed for power = .80 and alpha = .05. Nine subjects per group was estimated to yield enough power to find a true difference if one really exists. To obtain this estimate, an effect size was calculated from a study of Binder and Willis (1991) and found to be .70. This effect size is considered large and indicated a need for approximately 9 subjects per group. To be conservative, 15 subjects per group was used.

Suspected Malingerers

Suspected malingerers were obtained from the archival files of a private practice in Baton Rouge. The approximate 2300 total archival files were comprised of therapy cases (41%), vocational rehabilitation evaluations (18%), forensic neuropsychological evaluations (17%), medically-referred neuropsychological or psychological evaluations (15%), psychoeducational evaluations (6%), as well as other source referrals (3%). All forensic neuropsychological assessments were examined and those patients that met criteria for *possible*, *probable*, or *definite* malingering were included in the study. Of 396 forensic cases, 30 suspected malingerers were found yielding an estimated base rate of malingering as 8%.

All thirty participants were administered a traditional fixed neuropsychological battery based on the Halstead-Reitan Battery, along with effort measures including 1 or more of the following: TOMM, DCT, MFIT, PAI, and MMPI. The sample was randomly separated into two groups to allow for cross validation. The first sample consisted of 12 males and 3 females of the

following ethnicities: 8 Caucasians, 5 African-Americans, 1 Pacific Islander, and 1 of unreported origin. The mean age was 38.3 years ($SD = 12.9$) and mean level of educational attainment was 11.2 years ($SD = 1.6$). The second sample consisted of 12 males and 3 females of the following ethnicities: 7 Caucasians, 7 African-Americans, and 1 of unreported origin. The mean age was 37.0 years ($SD= 10.2$) and mean level of educational attainment was 10.6 years ($SD=2.5$).

Demographically Unmatched Simulated Malingerers

Thirty unmatched simulated malingerers were recruited from undergraduate psychology classes at Louisiana State University. This sample had a disproportionately high amount of females, therefore an additional 18 males were recruited for the study, yielding a total of 48 subjects. Participation was on a volunteer basis and 2 extra credit points were awarded. All participants were screened to ensure that they had no previous moderate to severe head injury, no neurological disease, no current psychiatric disorder, and were 18 years or older.

The sample was randomly separated into two groups to allow for cross validation. The first sample consisted of 9 males and 15 females of the following self-reported ethnicities: 21 Caucasians, 1 African-American, 1 Asian, and 1 Hispanic. The mean age was 20.1 years ($SD = 1.7$) and mean level of educational attainment was 13.3 years ($SD = 1.3$). The second sample consisted of 13 males and 11 females of the following self-reported ethnicities: 19 Caucasians, 2 African-Americans, 1 Asian, 1 Hispanic, and 1 of unreported origin. The mean age was 19.6 years ($SD = 1.3$) and the mean level of educational attainment was 13.3 years ($SD = 1.5$).

<u>Demographically Matched Simulated Malingerers</u>

A convenience sample of 31 simulated malingerers was recruited via newspaper ads and fliers from various outlets in the New Orleans and Baton Rouge areas. Three participants were excluded; one for history of a moderate head injury, one for the presence of a psychiatric disorder, and one subject withdrew before testing was complete, yielding a total of 28 participants. Participants were matched to archival malingerers on variables of age and race. Although attempts were made to match on education, the simulated malingerers' level of education was significantly higher than that of archival malingerers, preventing a match on this variable. Participation was on a volunteer basis and participants were entered into a lottery to win $300. All participants were screened to ensure that they had no previous moderate to severe head injury, no neurological disease, no current psychiatric disorder, and were 18 years or older.

The sample consisted of 9 males and 19 females of the following self-reported ethnicities: 21 Caucasians, 5 African-Americans, 1 Hispanic, and 1 of unreported origin. The mean age was 32.3 years ($SD = 11.4$) and the mean level of educational attainment was 13.9 years ($SD = 1.6$). For a comparison of demographic information see Table 2.

<div align="center">Materials</div>

Written informed consent was obtained from all participants prior to their inclusion in the study. The following tests were administered:

Table 2

Mean (SD) Demographic Information of Suspected and Simulated Malingerers.

	Age	Education
Sample 1		
Suspected	38.3 (12.9)	11.2 (1.6)
Simulated (Unmatched)	20.1 (1.8)	13.3 (*1.3*)
Sample 2		
Suspected	37.0 (10.2)	10.6 (2.5)
Simulated (Unmatched)	19.6 (1.3)	13.3 (1.5)
Total		
Suspected	37.7 (11.4)	10.9 (2.1)
Simulated (Matched)	32.3 (11.4)	13.9 (1.6)
Simulated (Unmatched)	19.9 (1.5)	13.3 (1.4)

All numbers are reported in years.

Structured Interview

A structured interview was developed to obtain the following information from participants: age, gender, race, education, neurological history, history of head injury, and current psychological status. This measure was administered only to simulated malingerers. Archival files of suspected malingerers were used to determine the same demographic information. See Appendix A.

Test of Memory Malingering (TOMM)

The TOMM is an effort measure which consists of two learning trials and a retention trial. Each learning trial is divided into two parts, a study phase and a test phase. The study phase contains 50 pictures presented one at a time for three seconds. Immediately following is the test phase where the participant must decide from two possible choices which picture he has studied. Following a fifteen- minute delay, the retention trial is administered. The retention trial consists of the test phase only. A score of one point is credited for every correct answer. The TOMM was administered according to standard instructions (Tombaugh, 1996). According to the test manual, those individuals who scored 50 out of 50 on Trials 1 & 2, are not required to complete the retention trial (Tombaugh, 1996). Therefore, a score of 50 out of 50 was given to those participants who, at the examiner's discretion, did not complete this trial.

Memory for Fifteen Items Test (MFIT)

The MFIT consists of fifteen items arranged on one page in three columns by five

rows. Participants are shown this page for 10 seconds and then are asked to draw the page

from memory. One point is awarded for each item correctly reproduced. See Appendix B.

Dot Counting Test (DCT)

The DCT consists of twelve index cards printed with either grouped or ungrouped

dots. The participant is asked to count the dots as quickly as possible. The total number

of errors is calculated. See Appendix C.

Personality Assessment Inventory (PAI)

The PAI consists of 344 statements on which a participant can answer as either

False, *Slightly True*, *Mainly True*, or *Very True*.   The completed form is entered into a

computerized scoring program which provides a print-out containing scores on 4 validity

scales and 9 clinical scales.

Wide Range Achievement Test- Third Edition (WRAT-3)

The reading subtest of WRAT-3 was administered to all participants to ensure at

least a fourth grade reading level as required by the PAI.

Subject Rating Scale

A subject rating scale, taken directly from Tombaugh (1996), asked the questions:

How successful do you think you were in your attempt to portray someone with a brain-

injury? How hard did you try? Subjects rated their answers on a 6 point Likert scale. This

measure was administered to control for those participants who reportedly did not try. See

Appendix D.

<div align="center">Design and Procedure</div>

Simulated Malingerers

All participants were given informed consent prior to participation and were

informed of the confidentiality of their responses to test items and questionnaires.

Participants were assigned an identification number to maintain anonymity of all

responses. After informed consent was obtained, participants were interviewed according

to the Structured Interview designed for this study. Participants that met exclusion criteria

were thanked and dismissed. The remaining participants were given the WRAT-3 to

obtain a reading level. Following the WRAT-3, participants were read a set of instructions

taken directly from a study by Tombaugh (1996). Instructions requested that the

participant perform as if he/she had experienced a head injury in a car accident and were

coached by their attorney to perform with demonstrable brain damage. Exact instructions

are as follows:

> In this study you will be asked to complete a set of tasks that are often used
> to measure a variety of changes that occur in people who have brain
> damage. As you take each test, we would like you to assume the role of
> someone who has experienced some brain damage from a car accident.
>   Pretend that you were involved in a head-on collision. You hit your
> head against the windshield and were unconscious for 15 minutes. You
> were hospitalized overnight for observation and then released. Gradually,
> over the past few months, you have started to feel normal again. However,
> your lawyer has informed you that you may get a larger settlement from the
> court if you look like you are still suffering from brain damage. In the real
> world, the usual purpose of the tests you are about to take is to determine if

the accident has produced any impairments in your abilities due to brain damage.

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

Immediately following the instructions, participants were administered the two learning trials of the TOMM followed by the MFIT and DCT. After a fifteen-minute delay the TOMM retention trial was administered followed by the PAI and Subject Rating Scale. Upon completion of these measures, participants were awarded with compensation and dismissed. Results of the assessment were then calculated and each participant was classified into either no category, or *possible, probable*, or *definite* categories. There was no external incentive contingent upon test performance in this sample; however, for uniformity across groups, classification was adjusted so that all simulated participants met criterion A.

Results

<u>Analyses of Hypotheses</u>

<u>Hypothesis 1a</u>

To test hypothesis 1a, that the frequencies of *possible, probable*, and *definite* classifications are significantly different in suspected and simulated malingerers, a chi-square analysis was used. Significance was considered at the *p* ≤.05 level. The independent variable was group membership (suspected vs. simulated) and the dependent variable was Slick et al. classification category (*possible, probable*, or *definite*).

The chi-square analysis was performed on the first random sample of archival and demographically unmatched malingerers (N = 39). Results indicate a near significant difference in the frequencies of classification category, *X²*= 7.52, *df* = 3, *p* = .057. In the sample of archival malingerers (n = 15), 1 patient was classified as *definite*, 9 patients were classified as *probable*, and 5 patients were classified as *possible* malingerers.   The modal classification category of archival malingerers was *probable*. In the sample of demographically unmatched simulators (n = 24), 1 participant met no classification category, 6 participants were classified as *definite*, 16 were classified as *probable*, and 1 was classified as *possible* malingering. The modal classification category of simulated malingerers was *probable* . The effect size was considered moderate, ɸ = .439. See Figure 2.

Figure 2

Frequency of Classification Category in Suspected and Simulated Malingerers (Sample 1):
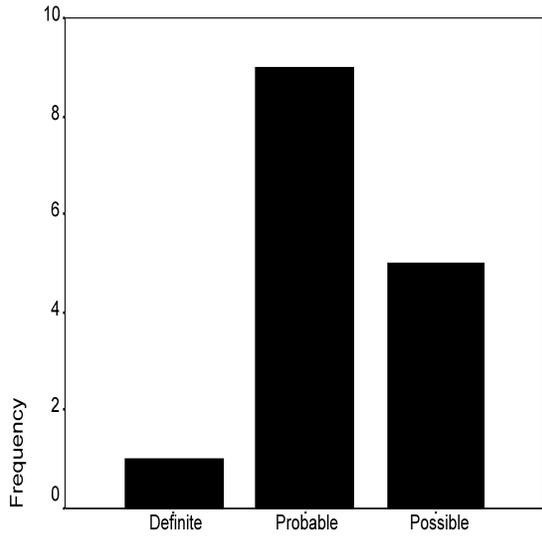
Slick Classification of Suspected Malingerers
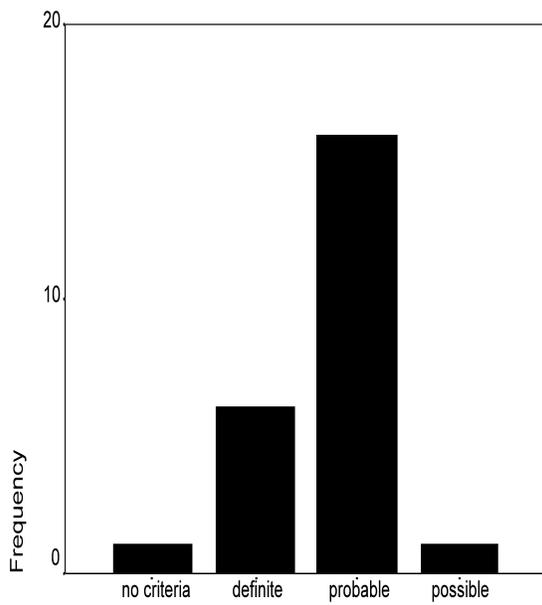
Slick Classification of Simulated Malingerers

To test hypotheses 1b-1d, that simulated malingerers perform significantly worse on the TOMM, MFIT, and DCT, compared to suspected malingerers, a 1 x 3 between subjects MANOVA was proposed. However, once data was collected it was revealed that there were not enough archival participants who had completed all three measures to allow for this analysis (9 subjects were needed for power = .80, alpha = .05), therefore hypotheses were tested using multiple independent t-tests. The p-value was changed from .05 to .01 to account for the increased error rate associated with multiple significance testing. For each t-test the independent variable was group classification (suspected vs. simulated) and the dependent variable was test performance (TOMM1, TOMM2, TOMMR, MFIT, or DCT).

Multiple independent t-tests were performed with the first random sample of archival (n = 15) and demographically unmatched simulated malingerers (n = 24). On the TOMM Trial 1, simulated malingerers ($M = 27.5$, $SD = 8.23$) performed nearly significantly worse than suspected malingerers ($M = 37.13$, $SD = 12.29$), $t(30) = 2.53$, $p = .017$. On the TOMM Trial 2, simulated malingerers ($M = 26.83$, $SD = 10.75$) performed significantly worse than suspected malingerers ($M = 39.25$, $SD = 10.77$), $t(30) = 2.83$, $p < .01$. On the TOMM retention trial, simulated malingerers ($M = 24.83$, $SD = 9.96$) performed significantly worse than suspected malingerers ($M = 37.13$, $SD = 24.83$), $t(30) = 2.93$, $p < .01$. However, on the MFIT, simulated malingerers ($M = 12.04$, $SD = 3.04$) did not perform significantly worse than suspected malingerers ($M = 9.31$, $SD = 3.95$), $t(35) =$

-2.35, $p = .025$. Similarly, on the DCT, simulated malingerers ($M = 6.46$, $SD = 3.45$) did not perform significantly worse than suspected malingerers ($M = 5.40$, $SD = 3.71$), $t (27) = -.62$, $p = .58$. See Table 3.

Hypothesis 2

To test hypothesis 2, that the results of hypothesis 1 will repeat in cross-validation, a chi-square analysis was performed in a second sample. Significance was considered at the $p \leq .05$ level. The independent variable was group membership (suspected vs. simulated) and the dependent variable was Slick et al. classification category (*possible, probable*, or *definite*).

A chi-square analysis was performed on the second random sample of archival and demographically unmatched malingerers (N = 39). Results indicated a significant difference in the frequencies of classification category, $X^2 = 11.67$, $df = 2$, $p < .05$. In the sample of archival malingerers (n = 15), no patients were classified as *definite*, 7 patients were classified as *probable*, and 8 patients were classified as *possible* malingerers. The modal classification category of archival malingerers was *possible*. In the sample of demographically unmatched simulators (n = 24), 6 participants were classified as *definite*, 16 were classified as *probable*, and 2 were classified as *possible* malingering. The modal classification category of simulated malingerers was *probable*. The effect size was considered moderate, $\phi = .547$. See Figure 3.

Table 3

Independent Sample t-tests for Suspected vs. Demographically Unmatched Simulators

	t	df	significance
TOMM Trial 1	2.53	30	.017
TOMM Trial 2	2.83	30	.008**
TOMM Retention	2.93	30	.006**
DCT	-.62	27	.581
MFIT	-2.35	35	.025

** $p < .01$

Figure 3

Frequency of Classification Category in Suspected and Simulated Malingerers (Sample 2):
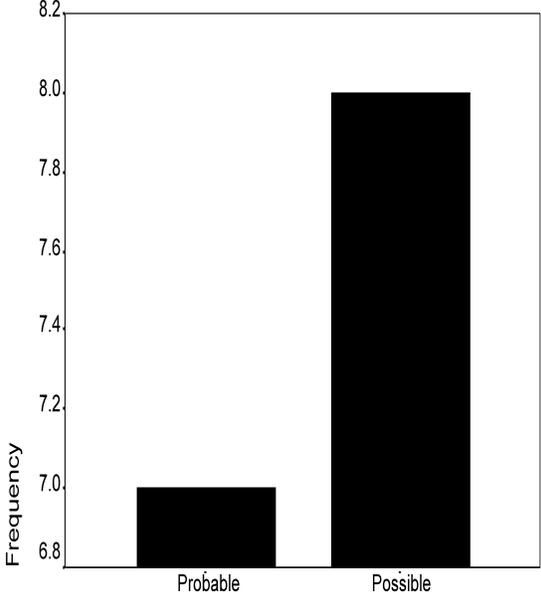
Slick Classification of Suspected Malingerers

Slick Classification of Simulated Malingerers

Hypothesis 3

To test hypothesis 3, that classification categories are not significantly different for demographically matched and unmatched simulators, a chi-square analysis was performed. Significance was considered at the $p \leq .05$ level. The independent variable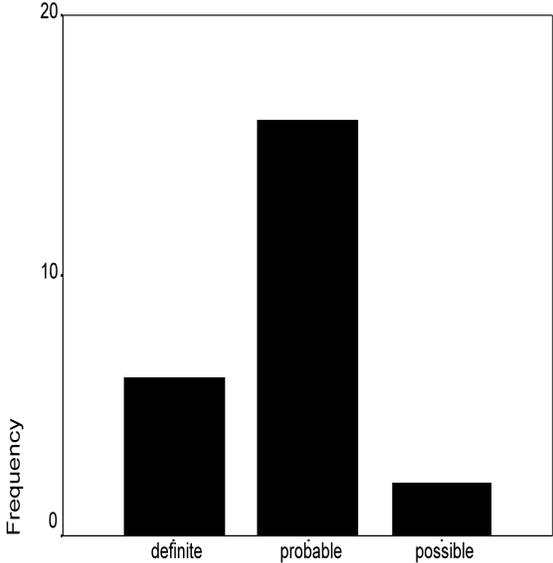 was group membership (demographically matched vs. demographically unmatched) and the dependent variable was Slick et al. classification category (*possible, probable*, or *definite*).

Demographic Analysis

To verify that demographically matched simulators were indeed matched to suspected malingerers, a independent samples t-test was run on variables of age, race, gender, and years of education. Results indicated no significant differences between matched simulators and suspected malingerers on variables of age, $t(56) = 1.77, p = .08$, and race, $t(56) = 1.54, p = .128$. The two groups were significantly different on years of education $t(52) = -.6.10, p < .05$ and gender $t(56) = -4.127, p < .05$. See Table 4.

Chi-Square Analysis

Results from the chi-square analysis indicated that there was no significant difference in the frequencies of *possible, probable*, and *definite* malingering in demographically matched and unmatched simulator samples, $X^2 = .269, df = 3, p = .966$ (N = 76). In the sample of demographically unmatched simulators (n = 48), 1 participant met no classification category, 12 participants were classified as *definite*, 32 were classified as *probable*, and 3 were classified as *possible* malingering. The modal classification

29

Table 4

**<u>Independent Samples t-test for Demographic Variables in Suspected and
Demographically Matched Simulators</u>**

	t	df	significance
Age	1.77	56	.082
Education	-6.10	52	.000**
Gender	-4.11	54	.000**
Race	1.55	56	.128

** $p < .01$

category for demographically unmatched simulators was *probable*. In the sample of

demographically matched simulators (n = 28), 1 participant met no classification category,

6 participants were classified as *definite*, 19 were classified as *probable*, and 2 were

classified as *possible* malingering. The modal classification category of demographically

matched simulators was *probable*. The effect size was considered small, ϕ = .06, likely

due to homogeneity between groups. See Figure 4.

Figure 4

<u>Frequency of Classification Category in Matched and Unmatched Simulated Malingerers:</u>
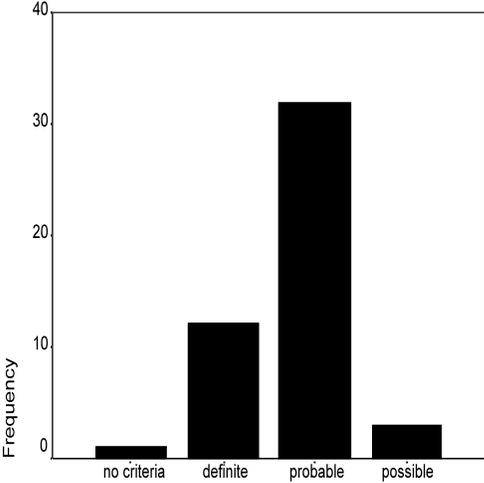
Slick Classification of Unmatched Simulators
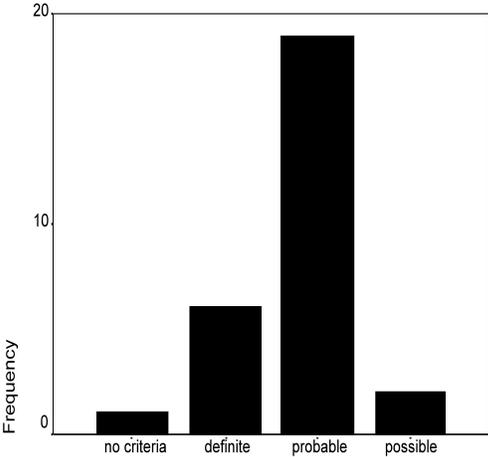
Slick Classification of Matched Simulators

Discussion

The majority of malingering research incorporates either the simulation design or the differential prevalence design. Both designs have been criticized for methodological flaws. The simulation design is criticized for external validity concerns, specifically the unknown generalizability of simulated malingerers to actual malingerers. The differential prevalence design is criticized for internal validity concerns, particularly for the ambiguity in which subjects are chosen. The present study attempted to address both internal and external validity concerns.

In the current study, both designs were used. This design was considered internally valid due to the strict criteria from which participants were chosen. Rather than selecting "at-risk" participants, archival files were culled and only individuals that met clearly specified criteria were selected for the study. The criteria, provided by Slick et al. (1999), are beginning to receive support from the research community and are believed to be a solid base from which to detect malingering (Mathias, Greve, Bianchini, Houston, & Crouch, 2002; Greve et al., 2002). This design, of only *possible, probable*, or *definite* malingerers, was compared to the typical simulation design, utilizing university undergraduates. The results demonstrate that simulated malingerers are significantly different from suspected malingerers. More specifically, the frequencies of *possible, probable*, and *definite* classifications were different. While this finding might lead one to conclude that there is little generalizability of simulation designs to forensic populations, there is a caveat. The modal frequency category for the first sample of archival and simulated malingerers was *probable* malingering, and this trend was nearly repeated in the cross-validation. This

finding may suggest that external validity can be preserved within a simulation design if only *probable* malingerers are used.

Further comparisons of simulated and suspected malingerers were made on specific test performance. Overall, the results point to simulated malingerers performing significantly worse than suspected malingerers. This was observed on two trials of the TOMM. While performance was significantly lower in simulators, this difference was not considered significant in terms of test interpretation. The mean score for both groups fell below the cut-off score for a probable response bias and above the cut-off score for a negative response bias, therefore, both groups would have been classified the same way in clinical practice. While no significant difference between groups was found, a similar relationship was observed on the MFIT. The mean performance for both groups was above the cut-off score of 7, so neither group performance would have been interpreted as a failure. However, if the cut-off score was adjusted to 9 (as is often used in clinical practice), there would be a significant difference in terms of test interpretation. The mean archival group performance would have been interpreted as a marginal failure, while the mean simulation group performance would have been interpreted as a pass.

There was no significant difference in test performance observed on the DCT. Both the simulators and suspected malingerers performed similarly and both mean performances were interpreted as a failure. The reason no significant difference was observed is most likely due to the small number of archival patients who had completed this measure. While

a sample size of 9, for power = .80 and alpha = .05, was required, only 5 archival subjects had completed a DCT, therefore limiting the ability to find a true difference.

The second purpose of this study was to determine whether demographic matching of simulating malingers is necessary. This study compared demographically matched simulators to demographically unmatched simulators. The results of this comparison indicate that there is no significant difference between matched and unmatched simulators on the frequencies of *possible, probable*, and *definite* classification. This finding, however, should be interpreted with caution. Matched simulators were matched to archival malingerers on only two variables, age and race. While it appears that matching on these two variables does not make a difference, perhaps matching on other variables do. Before a clear determination can be made regarding this hypothesis, future research is needed.

The overall results of this study showed, first, it does not appear that simulated malingerers adequately represent actual malingerers in either classification category or test performance. However, research utilizing the simulation design should not be considered fruitless. It appears that the majority of simulating malingerers represent the majority of suspected malingerers, therefore supporting the usefulness of this design. It is only for extreme performances (*possible* and *definite*) that a simulation design should be interpreted cautiously. Furthermore, while simulated malingerers perform significantly worse than suspected malingerers on a variety of effort measures, these differences are not large enough to influence the clinical interpretation of the performance. Lastly, it appears that

matching simulated malingerers to archival malingerers on variables of age and race is not necessary.

There are several limitations to this study. In regard to suspected malingerers, there was a large amount of variation in the number and type of effort measures administered as part of the original assessment. Consistency across the test battery may provide for more meaningful analysis of this sample. In addition, the base rate of malingering in this clinical sample (8%) is lower than that reported in other forensic practices, perhaps skewing the data. Future research can address these concerns by studying suspected malingerers obtained from forensic practices with higher base rates, as well as using archival data with more consistency among the test battery. As for the demographically matched sample, the simulators were only matched on two demographic variables, thus limiting the interpretation of the findings. In order to adequately determine whether matching is necessary, matching should occur on several variables. Perhaps the next step to addressing this hypothesis would be to implement demographic matching on variables of education and socioeconomic status. This would allow for a better analysis of factors most likely to affect malingering sophistication.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bernard, L.C., & Fowler, W. (1990). Assessing the validity of memory complaints: performance on brain-damaged and normal individuals on Rey's task to detect malingering. *Journal of Clinical Psychology, 49*, 432- 436.

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*, *5*, 715-728.

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

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

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

Daubert v. Merrell Dow Pharmaceuticals, Inc. 509 U.S. 579, 113 S.Ct. 2786

Faust, D. & Nurcombe, B. (1989). Improving the accuracy of clinical judgement. *Psychiatry: journal for the study of interpersonal processes, 52*(2), 197-208.

Ford, C.V.(1983). *The somatizing disorders: illness as a way of life*. New York: Elsevier.

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

Frederick, R.I., Sarfarty, S.D., Johnston, J.D., & Powel, J. (1994). Validation of detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology*, *8*, 118- 125.

Frederick, R. (2002). A review of Rey's strategies for detecting malingered neuropsychological impairment. *Journal of Forensic Neuropsychology*, *2*, 1-25.

Goldberg, J.O. & Miller, H.R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task assessing the validity of memory complaints. *Journal of Clinical Psychology*, *42*, 792-795.

Gouvier, W.D., Hayes, J., & Smiroldo, B. (1998). "The significance of base rates, test sensitivity, test specificity, and subject's knowledge of symptoms in assessing TBI sequelae and malingering." *Detection of Malingering during Head Injury Litigation*. New York: Plenum Press.

Gouvier, W.D. (1999). "Baserates and clinical decision making in neuropsychology". *Forensic Neuropsychology Fundamentals and Practice.* Royersford, PA: Swets & Zietlinger.

Gouvier, W.D., Lees-Haley, P., & Hammer, J. H. (2003). The neuropsychological examination in the problem of detecting malingering in the forensic arena; Costs and benefits. *Clinical Neuropsychology and Cost Outcomes Research: A beginning*. New York: The Psychology Press.

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

Greve, K., Bianchini, K., Mathias, C., Houston, R., & Crouch, J. (2002). Detecting malingered performance with the Wisconsin Card Sorting Test: a preliminary investigation in traumatic brain injury. *The Clinical Neuropsychologist*, *16* (2), 179-191.

Groth-Marnat, G. (1997). *Handbook of Psychological Assessment*. (3rd ed.). New York: John Wiley & Sons, Inc.

Haines, M.E., & Norris, M.P., (1995). Detecting the malingering of cognitive deficits: an update. *Neuropsychology Review*, *5*(2), 125-148.

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

Lee, G.P., Loring, D.W., & Martin, R.C. (1992). Rey's 15 visual memory test for the detection of malingering" normative observations on patients with neurological disorders. *Psychological Assessment*, *4*, 43-46.

Lezak, M.D. (1983). *Neuropsychological assessment*. (2nd ed.). New York: Oxford Press

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

Mathias, C., Greve, K., Bianchini, K., Houston, R., & Crouch, J. (2002). Detecting malingered neurocognitive dysfunction using the reliable digit span in traumatic brain injury. *Assessment*, *9* (3), 301-308.

Meehl, P., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*, 194-216.

Millis, S. & Kler, S. (1995). Limitations of the Rey-Fifteen Item Test in the detection of malingering. *Clinical Neuropsychologist*, *9*, 241-244.

Morey, L. (2003). Essentials of PAI Assessment. New York: John Wiley & Sons, Inc.

Reynolds, C (1998). "Common sense, clinicians, and actuarialism.*"Detection of Malingering during Head Injury Litigation*. New York: Plenum Press.

Rogers, R., Bagby, R.M., & Dickens, S.E.(1992). *Structured Interview of Reported Symptoms: Professional Manual*, Oddess, FL: Psychological Assessment Resources.

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

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

Shadish, W., Cook, T., & Campbell, D. (2002) *Experimental and quasi-experimental designs for generalized causal inference.* New York: Houghton Mifflin Company

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

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

Tombaugh, T. (2002). The Test of Memory Malingering (TOMM) in forensic psychology. *Journal of Forensic Neuropsychology*, *2* , 66-96.

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

Vallashajosula, B. & van Gorp, W.G. (2001). Post-Daubert admissibility of scientific evidence on malingering of cognitive deficits. *The Journal of the American Academy of Psychiatry and Law*, *29*, 201-215.

Structured Clinical Interview

Subject #:_____   Examiner:_____

Age: _____

                                                        Matched   Unmatched

Race: _____

Gender: _____

Highest grade completed: _____

Do you currently, or have you previously had any type of neurological disorder, for example epilepsy? If so please explain _____
_____
_____
_____

Have you every been hit on the head so hard that you blacked out? If so please explain when and how long you were unconscious._____
_._____
_____
_____

Are you being treated for a psychological disorder? If so please explain._____
_____
_____
_____

Appendix B

Memory for Fifteen Items Test

Subject #:_____   Examiner:_____

Column 1 Correct: _____

Column 2 Correct: _____

Column 3 Correct: _____

Column 4 Correct: _____

Column 5 Correct: _____

Total Correct : _____

Appendix C

Dot Counting Test

Subject #:_____   Examiner:_____

(Circle One)

Card 1:   Error/No Error

Card 2:   Error/No Error

Card 3:   Error/No Error

Card 4:   Error/No Error

Card 5:   Error/No Error

Card 6:   Error/No Error

Card 7:   Error/No Error

Card 8:   Error/No Error

Card 9:   Error/No Error

Card 10:   Error/No Error

Card 11:   Error/No Error

Card 12:   Error/No Error

Total Number of Errors: _____

Appendix D

Subject Rating Scale

Subject #:_____   Examiner: _____

1.) How successful do you think you were in your attempt to portray someone with a
      brain-injury? (Circle one)

                     Not at all   Very

                     0   1   2   3   4   5

2.) How hard did you try? (Circle one)

                     Not at all   Very

                     0   1   2   3   4   5

Consent Form

Louisiana State University
236 Audubon Hall
Baton Rouge, LA 70803-5501
(225) 578-1494 Phone - (225)578-4661 Fax
_____

1. Study Title:
     A Profile and Comparison of Simulated and Suspected Malingerers

2. Performance Site:
     Louisiana State University

3. Investigators:
     The investigators listed below are available to answer questions about the research,
     M - F, 8:00 a.m. - 4:00 p.m.

     Wm. Drew Gouvier, Ph.D. & Adrianne Brennan, B.A.
     (225) 578-1494

4. Purpose of the Study:
     The purpose of this research is to determine whether people asked to malinger
     perform in the same way as true malingerers.

5. Subjects:
     A. Inclusion criteria: $\geq$ 18 years old
                                  Current undergraduates at LSU

     B. Exclusion criteria: Individuals who have suffered a moderate or severe
                                  head injury
                                  Neurological disease or seizure disorder
                                  Present psychological disorder

     C. Maximum number of subjects: 60

6. Study Procedures:
Each subject will be interviewed about their medical and psychological history and take five tests on which they will be asked to perform as if they had a head-injury. Interview plus test administration should not exceed two hours and will occur at one scheduled appointment.

7. Benefits:
Each undergraduate subject will receive two (2) extra credit points for full participation in this two (2) hour study. Other participants will be entered into a lottery to win $300. Information gained from this study may help us to better understand and improve current psychological research in the area of malingering.

8. Risks/Discomforts:
There is no known risk associated with participation in this study above what might be experienced in an average day.

9. Injury/Illness:
To assure that subject's privacy is respected, this study will be anonymous.

10. Right to Refuse:
Participation in this study is completely voluntary and subjects may change their minds and withdraw from the study at any time without penalty.

11. Privacy:
Subjects' names on consent forms will not be able to be linked to interview and questionnaire responses. Additionally, consent forms will be stored separately from data.

The LSU Institutional Review Board (which oversees university research with human subjects) and Wm. Drew Gouvier, Ph.D. may inspect and/or copy the study records.

Results of this study may be published, but no names or identifying information will be included in the publication.

12. Financial Information:

There is no cost to the subjects. Subjects will receive two (2) extra credit points, or a be entered into a lottery to win $300.

13. Withdrawal:
    You may withdraw from this study at any time, however, extra credit points or lottery entry will not be given for less than full participation. To withdraw, inform the principle investigator or research assistant of your decision.

14. Removal:
    If it becomes apparent that the subject is not responding in a forthright manner or additional information suggesting that a subject meets exclusion criteria is disclosed later in the study, the subject will be removed from the study without his or her consent.

The study has been discussed with me and all my questions have been answered. I may direct additional questions regarding study specifics to the investigator or research assistants. If I have questions about subjects' rights or other concerns, I can contact Robert C. Matthews, Chairman, LSU Institutional Review Board, (225) 578-8692. I agree to participate in the study described above and acknowledge the investigator's obligation to provide me with a signed copy of the consent form.

Subject Signature _____

Subject Name (Print) _____

Date _____

Witness Signature _____

Date _____

Vita

Adrianne Brennan was born in 1978 in New Orleans, Louisiana. She received a Bachelor of Arts Degree with Honors in psychology from the University of New Orleans in 2001. In that same year, she was awarded the Distinguished Undergraduate Student Award and the Richard D. Olsen Award in Experimental Design and Methodology from the Psychology Department of the University of New Orleans. In 2003, she began a doctoral program in clinical psychology at Louisiana State University. She currently works under Dr. Wm. Drew Gouvier and is specializing in neuropsychology.