2010-06-25

Malingering Detection among Accommodation-Seeking University Students

Spencer Paul Clayton
Brigham Young University - Provo

Malingering Detection among Accommodation-Seeking

University Students

Spencer Clayton

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Bruce N. Carpenter, Chair
Michael Larson, PhD
Patrick Steffen, PhD
Niwako Yamawaki, PhD
Ross Flom, PhD

Department of Psychology

Brigham Young University

August 2011

ABSTRACT

Malingering Detection among Accommodation-Seeking

University Students

Spencer Clayton

Department of Psychology

Doctor of Philosophy

Universities have increasingly sought to provide accommodative services to students with learning disorders and Attention-Deficit/Hyperactivity Disorder (ADHD) in recent decades thereby creating a need for diagnostic batteries designed to evaluate cognitive abilities relevant to academic performance. Given that accommodative services (extended time on tests, alternate test forms, etc.) provide incentive to distort impairment steps should be taken to estimate the rate at which students distort impairment and to evaluate the accuracy with which symptom distortion is identified. In order to address these concerns, the Word-Memory Test, Test of Memory Malingering, and Fake Bad Scale (of the MMPI-2) were compared in terms of their clinical utility in a university sample within a two-part study. In the first portion of the study, an analogue design (which included a control group (n = 29) and an experimental group (n = 30) that was asked to simulate an academic disability) was used to calculate the sensitivity and specificity of each measure. In the second portion of this study, scores were collected for 121 consecutively presenting students who were evaluated for academic difficulty at a large private university. Failure rates on measures of malingering placed the base rate of malingering within this population between 10 and 25 percent. The Word-Memory Test (WMT) demonstrated the most robust sensitivity and specificity. The modest sensitivity of the Test of Memory Malingering (TOMM) can be partially explained by the ease with which the measure is completed by university students as well as the format of its presentation. Although the scores on Fake Bad Scale (FBS) are modestly correlated with group membership (between controls and simulators), its use should be discouraged in this context due to poor sensitivity and to high rates of false positives.

Keywords: malingering, university students, Fake Bad Scale, TOMM, WMT

ACKNOWLEDGEMENTS

I am grateful to my committee, especially my chair Bruce N. Carpenter, for their support and willingness to accommodate my schedule. I offer a special thanks to Michael Brooks of the University Accessibility Center. Without his efforts this project would not have been possible. Finally, I would like to thank my wife for her patience and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Malingering Detection among Accommodation-Seeking University Students

The need to detect malingering is ever present in contexts that require mental health professionals to draw conclusions that will ultimately impact a person's access to resources. These resources can act as powerful motivations to distort, even unconsciously, the self-presentation of symptomatology on which special access is contingent. This is also true in university settings where decisions must be made regarding accommodations offered to students based on the outcomes of psycho-educational evaluations. A simple search of the literature will demonstrate that, while there are literally hundreds of articles relating to university students with learning disorders or Attention-Deficit/Hyperactivity Disorder (ADHD), there have been almost none regarding malingering in this context until recent years (Frazier, Frazier, Busch, Kerwood, & Demaree, 2008; Harrison, Edwards, & Parker, 2007; Quinn, 2003; Sullivan, May, & Galbally, 2007). Because accommodative resources are limited and these benefits provide incentive for symptom distortion, and because unjustified accommodations impact the accuracy of the evaluative procedures of the educational system, good practice in this area should perhaps include methods for malingering detection.

Although it is assumed that malingering detection methods developed in other contexts may be relevant and effective here, the lack of empirical evidence indicates a need to evaluate that assumption. Such research should, for example, examine the base rate of malingering within this group and the sensitivity and specificity of malingering measures. This study attempted to address these issues by evaluating the performance of three commonly used measures in both an analogue and a clinical sample. The findings of this study may assist clinicians in improving classification rates of their testing batteries, and aid administrators in the allocation of resources to individuals who genuinely need them.

Disabilities and Accommodations

Legislation requires universities to provide reasonable accommodations to individuals with disabilities including person's diagnosed with a learning disorder or ADHD (Latham, 2005). Accommodations for these diagnoses might include text books on tape, note takers, extended time on tests, private testing rooms, and/or alternate test forms. Also, students who have been awarded accommodations during their university education may go on to receive similar accommodations on standardized exams such as the GRE, LSAT, or state licensing exams, and may develop expectations for similar accommodations in the workplace. Clearly some individuals will perceive these benefits as incentive to feign or exaggerate symptoms during an evaluation given that these accommodations are typically contingent upon a diagnosable impairment. Further, given that we cannot remove all classification errors from our evaluations, policy makers may need to consider the values attached to avoidance of false positive versus false negative errors (i.e., whether priority should be given avoidance of erroneously denying legally disabled persons appropriate accommodations or avoidance of awarding limited resources to individuals feigning impairment: maximize specificity versus sensitivity).

The typical justifications for accommodations are diagnosable conditions which either impact learning or impact demonstrations of learning (e.g., performance on examinations) in circumscribed ways that can be largely overcome through reasonable accommodations. The two primary types of conditions historically viewed as meeting this definition are learning disorders and ADHD. Learning disorders are defined as a discrepancy between achievement in math, reading, or writing and a person's expected level performance given their age, intelligence, and education. ADHD is defined by symptoms in one or more of three categories: inattention, hyperactivity, and impulsiveness (American Psychological Association, 2000). These disorders are commonly assessed with individually administered measures of intelligence, achievement and attention. Also, testing batteries may include interviews, review of educational and other histories, omnibus measures of personality to screen for alternative, impairing psychiatric conditions, and tests of malingering to screen for symptom distortion. Even when good data is collected, variability in diagnosis can occur due to clinicians' tacit differences in thresholds for determining the presence of symptoms required for diagnosis. For example, with learning disability the discrepancy required for diagnosis, even when quantified, is known to be set at different levels (e.g., 1.5 standard deviations versus 2) across different settings. There are many potential sources of variability that contribute to diagnostic outcomes, including the willful production or distortion of symptoms. Interestingly, a recent study found that effort had a greater effect on the outcomes of psychological measures than did brain injury or illness (Green, Rohling, Lees-Haley, & Allen, 2001) arguing for the need to assess for malingering as it appears to be one of the most important sources of variability.

Malingering

Malingering is traditionally defined as the deliberate fabrication or exaggeration of symptoms for secondary gain (American Psychological Association, 2000) and a diagnosis of malingering sometimes implies underlying psychopathology. However, it has also been conceptualized as an adaptive response in an adversarial process (Rogers, 1990; Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998). This latter approach suggests that malingering is more a function of situations rather than of persons (such as a more general willingness to lie). Further, many conceptualizations recognize that self-awareness of malingering is on a continuum, calling into question whether deliberateness is an important element. Whether malingering is conceptualized as adaptive or pathological, clinicians are continually called upon to identify it due to problems associated with misclassification. In a survey conducted by Mittenberg, Patton, Canyock, and Condit (2002), clinical neuropsychologists reported that the prevalence of malingering varied widely depending on the context of the evaluation. Specifically, clinicians reported that they encountered malingering on average in 29% of personal injury cases, 30% of disability cases, 19% of criminal cases, and 8% of medical cases. This finding of variable base rates across different settings has an important implication for assessments in university settings: namely, one cannot expect the base rate of malingering to match that of other samples.

Malingering Assessment

Two operationalizations of malingering seem to predominate among mental health professionals: feigning/exaggerating symptoms or denying ability (Rogers, Harrell, & Liff, 1993). In other words, feigned psychopathology functionally presents as the over-reporting of symptoms whereas feigned cognitive impairment functionally presents as underperforming. These two operationalizations have contributed to the development of tests such as structured interviews, self-report screening measures, subscales on omnibus measures of psychopathology, and most recently, memory tests employing graphemes or pictures of objects. In a survey of practicing neuropsychologists, of those cases in which malingering was diagnosed, that conclusion was supported by data from forced-choice measures 57% of the time and by validity scales from omnibus measures of personality 38% of the time (Mittenberg, Patton, Canyock, & Condit, 2002).

Measures intended to capture the feigning or exaggerating of symptoms primarily rely on the over-reporting of symptoms. For example, the F subscales (F for fake bad) on the Minnesota Multiphasic Personality Inventory are elevated when items representing symptoms that are rarely experienced together, even among severe populations, are endorsed with unusual frequency (Graham, 2006). Evidence suggests that, while this method is effective for detecting exaggerated psychiatric symptoms, these measures are relatively insensitive to neuropsychological symptoms, which are more ability-based (Gervais, Ben-Porath, Wygant, & Green, 2007; Larrabee, 1998, 2003; Ruocco, et al., 2008). Also, effort appears to account for approximately half the variance in measures of cognitive functioning in some settings (Green, et al., 2001). Thus, measures specifically designed to detect the denying of ability (e.g. feigning impairment on neuropsychological measures) contribute incremental value to malingering detection in the course of neuropsychological evaluations.

Measures intended to capture the denial of ability typically rely on the low face validity of the test. Specifically, they involve tasks that may appear difficult when in actuality they are relatively easy. This method facilitates discrimination between groups (Green, 2005) and entices malingerers to display impairment beyond that expected of anyone except those with extreme head injuries (Merten, Bossink, & Schmand, 2007). Historically, these measures depended on below chance scores (i.e. scores lower than those that would be achieved by guessing alone) to support a diagnosis of malingering, but this method led to poor sensitivity. More recently, test developers have improved sensitivity by creating cut-scores after generating and comparing distributions for malingers and non-malingerers or simulators and controls (Green, Lees-Haley, & Allen, 2002; Tombaugh, 1996).

The emergence of new methods for detecting feigned cognitive impairment led to a need for professional standards regarding malingering detection in this area. Slick, Sherman, and Iverson (1999) proposed that performing below chance on a forced choice measure be considered definite malingering while performing below normative cutoffs on at least two or more measures designed for detecting poor effort be considered probable malingering. However, others have suggested that below chance scores are not the only definitive indicators of malingering (Boone, 2007). Simply falling below the normative cutoffs on multiple measures designed for malingering detection seems to be relatively definitive (Larrabee, 2008).

Measure Development

Critics have pointed out that validation of malingering detection tools should include both analogue and known-groups designs (Rogers & Cruise, 1998; Rogers, et al., 1993). Typical analogue designs use university students simulating malingering. While concerns have been raised about the external validity of these studies due to student simulators being significantly harder to detect than actual malingerers (Haines & Norris, 2001), research supports the generalizability of such designs (Brennan & Gouvier, 2006), particularly when the sample is representative of the group of interest and simulators are given realistic scenarios (Rogers & Cruise, 1998). Also, analogue designs have the inherent advantage of allowing researchers to randomly assign subjects to groups and manipulate the variable of malingering.

Known-groups designs are less common than analogue designs, but they have the advantage of increased generalizability and the ability to shed light on base rates in a given population. However, there are limitations to this design as well, particularly when correlational statistics are the sole method of analysis. As Butcher, Gass, Cumella and Williams (2008) aptly stated, "one could argue that any sign of increased symptoms and lowered neuropsychological performance are the reasons for the compensation seeking, rather than the compensation seeking being the reason for the altered presentation" (p. 9). Also, several threats to internal validity are introduced by non-random group assignment (Kazdin, 2003). However, these limitations can be addressed through the use of replicable procedures. When known-groups designs are used in tandem with analogue designs, stronger conclusions can be made regarding test properties.

Sensitivity, Specificity, and Base Rates

While sensitivity and specificity are viewed as test characteristics that are stable across differing base rates, positive and negative predictive power are not (Dawes, 1962; Glaros & Kline, 1998; Rosenfeld, Sands, & Van Gorp, 2000). Positive predictive power is particularly relevant to malingering detection as it represents the likelihood that a person predicted to be malingering is actually malingering. Given that empirical data support the view that base rates vary widely across contexts, one can infer that the positive predictive power of a measure is unknown unless used in a context where the base rate is clearly established. Also, sensitivity and specificity are not necessarily generalizable across settings because tests may not perform equally well among differing populations. For example, because university students simulating malingering tend to be more difficult to detect than actual malingerers (Haines & Norris, 2001), tests may appear less sensitive when used with this group.

Malingering Detection with Psychological Measures

The Test of Memory Malingering (TOMM; Tombaugh, 1996) and the Word Memory Test (WMT; Green, 2005) are two examples of effort tests that have garnered attention in recent years due to their ease of administration and significant research base. Although they commonly assess the fabrication of memory impairments specifically, they have been found to be useful in the detecting of fabrication of cognitive impairment generally (Bauer, O'Bryant, Lynch, McCaffrey, & Fisher, 2007; Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005; Green, et al., 2002; Merten, et al., 2007). Also, because studies suggest that both the TOMM and WMT are insensitive to learning disabilities or ADHD in children (Green, 2005; Tombaugh, 1996), it can reasonably be inferred that students meeting admission criteria for a university will not fail either test simply due to such an impairment.

The TOMM has been found to have near perfect specificity (Gervais, Rohling, Green, & Ford, 2004; Rees, Tombaugh, Gansler, & Moczynski, 1998) due to its insensitivity to mild cognitive impairment. In fact, one researcher found that adults with mild mental retardation performed above the TOMM's published cut-offs for malingering (Simon, 2007). Interestingly, some studies have shown that a shortened form of the TOMM produces similar specificity (Bauer, et al., 2007; Gavett, O'Bryant, Fisher, & McCaffrey, 2005), suggesting that Trial 1 of the TOMM may serve as an effective screening measure. However, data suggests that this test may have modest sensitivity (Gervais et al., 2004; Weinborn, Orr, Woods, Conover, & Feix, 2003). The excellent specificity and modest sensitivity may be explained by the extreme ease with which this test is completed.

The WMT has received high praise from reviewers (Hartman, 2002) for its contribution to detection of poor effort in neuropsychological evaluations. This measure is reputed to have both excellent specificity and sensitivity (Green, 2005; Green, et al., 2002) suggesting that it is a valuable contribution to malingering assessment. It has been suggested that its excellent sensitivity can be attributed to low face validity due to the test's format (Sullivan, et al., 2007). Also, because the WMT uses the same word pairs across trials, it is possible to calculate scores for inconsistent responding giving the measure an added dimension of sensitivity. The WMT appears to be able to discriminate between poor effort and several types of genuine impairment.

The Minnesota Multiphasic Personality Inventory, 2nd Edition (MMPI-2) is the most widely used psychological test in the United States (Graham, 2006), and it is frequently used in neuropsychological evaluations. The Fake Bad Scale (FBS) was rationally derived by selecting 45 of the items from the 567 questions already existing on the MMPI-2 and was intended for detecting malingerers among personal injury claimants (Lees-Haley, et al., 1991). The FBS has received increasing attention as its proponents have asserted that it is more effective in detecting malingering of cognitive deficits relative to other MMPI-2 validity scales (Fox, Gerson, & Lees-Haley, 1995; Greiffenstein, Baker, & Gola, 2002; Larrabee, 2003). However, questions have been raised regarding its construct validity due to its tendency to be elevated in the presence of somatic problems and general maladjustment (Arbisi & Butcher, 2004; Butcher, Arbisi, Atlis, & McNulty, 2003; Butcher, Gass, Cumella, Kally, & Williams, 2008). The concern, then, is that the FSB, in valuing sensitivity over specificity, detects additional cases of malingering at the expense of added false positives. Despite these concerns, the FBS was added to the standard output of the MMPI-2 scoring program making it readily available to mental health professionals whether they are prepared to accurately interpret it or not (Armstrong, 2008). Therefore, studies are needed to clarify its potential efficacy across settings, including among accommodation-seeking university students, in order to avoid misinterpretation of FBS scores.

Symptom Exaggeration in LD and ADHD

Studies have refined the use of neuropsychological measures in a variety of contexts (Greiffenstein, Fox, & Lees-Haley, 2007). However, as stated above, little to nothing has been published regarding the efficacy of such tests among accommodation-seeking university students. Also, ADHD in particular is easily faked on self-report measures (Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008). Initial estimates of the base rate of malingering among accommodation seeking university students vary widely, from 8 % (Harrison, et al., 2007) to 46.7% (Sullivan, et al., 2007), due, in part, to varying methods of malingering assessment and the use of measures that have yet to be validated in this population. The relative absence of empirically supported means of malingering detection in this setting suggests that further study is needed.

Although justified, making accommodative services in university settings contingent upon receiving a diagnosed academic disability provides clear incentive for students to distort their impairment. Preliminary estimates of the base rate of malingering in this population support the suggestion that malingering assessment should be included in routine practice. Developments in the theory of malingering and in the methods for assessing it present several potential means by which malingering among accommodation-seeking university students may be approached. Intuitively, measures intended to capture the denying of ability seem a promising avenue. However, the widespread availability of the MMPI-2 and recent reports of the addition of the FBS to the scored report suggest that the use of measure should also be explored in this context.

Variables and Hypotheses

This study had two parts. Using an analogue design in which participants simulated accommodation-seeking college students, the sensitivity and specificity of three commonly used measures for detecting malingering was examined. During the analogue portion of the study, malingering was simulated by giving subjects instructions, a scenario, and monetary incentives to display symptomatic behavior. These independent variables were hoped to produce a significant change in group means on three measures of malingering between the malingering condition and the control group. These dependent variables are the Test of Memory and

Malingering (TOMM), The Word Memory Test (WMT), and the Fake Bad Scale (FBS) – a validity scale of the MMPI-2.

A second portion of this study attempted to establish the base rate of malingering among a clinical sample of accommodation-seeking university students by using cut-scores generated from the data collected during the analogue study. Also, failure rates for each test were compared to better understand how choice of index impacts malingering conclusions.

The principle hypotheses of the study are as follows. First, the TOMM would display excellent positive predictive power (a high percentage of those predicted to be malingerers will actually be simulating malingering), while displaying poor sensitivity (allowing a significant portion of malingerers to remain undetected) due to the extremely low difficulty of this task. Second, the WMT would have similar positive predictive power as the TOMM, but, due to its low face validity (appearing to be much more difficult than it actually is), it will be more sensitive. Third, the FBS would be poorly correlated with malingering due to problems with construct validity, and it will display poor positive predictive power (high numbers of false positives). The main hypothesis regarding the clinical sample was that the base rate of malingering will be relatively low (10 percent or lower) when compared to other settings, and that this low base would vitiate the clinical utility of the FBS. Additionally, exploratory analyses will be conducted to evaluate whether there is any added benefit to including more than one measure of malingering, and whether cut-scores could be altered to improve diagnostic accuracy within this population.

Study 1

Method

Participants. In order to approximate the university students examined in the clinical sample, participants for the analog portion of this study were recruited among students enrolled at the same university. Power analyses using effect sizes in similar malingering studies (Brennan & Gouvier, 2006) indicated that samples as small as 18 (9 per group) would likely produce large enough effect sizes to make significant group differences apparent (Cohen, 1988). However, because university students tend to be more difficult to detect when participating in analogue studies (Haines & Norris, 2001), a more conservative sample of 60 (30 students per block) was collected. Participants were randomly assigned to two blocks, one block instructed to malinger and the other instructed to complete three measures to the best of their abilities. The groups only approximated each other in terms of age and race, as these variables do not appear to have a significant impact on scores in analogue studies of malingering (Brennan & Gouvier, 2006). However, gender ratios reflected the ratio generally seen among students presenting for accommodations at the university at which this study is conducted (approximately .5).

Measures. Several measures of malingering were used in this study. Each employs a somewhat different approach to assess exaggeration or feigning of symptoms.

Test of Memory Malingering (TOMM). The TOMM (Tombaugh, 1996) is made up of visual stimuli that are presented in a serial manner. Subjects are then asked to identify pictures they have been shown previously in a dichotomous forced choice format. The entire test consists of two learning trials and a retention trial, and 50 pictures are presented in each learning trial. The retention trial is only presented if the participants' score below the recommended cut score for the second learning trial, and is presented 15 minutes after the learning trials. The total duration of the test ranges from 15 to 20 minutes, excepting the waiting period for the retention trial.

During the initial development of this test, the author employed an analogue design in which participants were asked to simulate cognitive impairment in the context of litigation after having sustained in injury, and after having been told that sustained brain damage would increase the settlement. Additionally, participants were informed that they would receive a cash prize of $50 if they presented the most convincing simulation. The average scores for trial one and two among the 27 simulators were 32.5 (*SD* = 7.5) and 35.3 (*SD* = 9.4) whereas the average scores for the 22 controls were 48.9 (*SD* = 1.6) and 49.9 (*SD* = 0.2) respectively (Tombaugh, 1996). More descriptions of test validity were provided above.

Word Memory Test (WMT). The WMT (Green, 2005) is primarily computer administered and involves the presentation of 20 word pairs. The original word pairs are then displayed in different formats with distracters. The test is divided into two portions (immediate recognition and delayed recognition) that must be administered approximately 30 minutes apart. The total duration of the two portions is approximately 25 minutes. The university students sampled met the basic requirements for completing this test as only a third grade reading level is required (Green & Flaro, 2003).

This measure was developed by the author primarily in the context of evaluations performed on brain-injured, compensation-seeking litigants, and it was found to account for more variance in test scores than did brain injury (Green et al., 2002). Additionally, analogue designs that included controls, instructed simulators, and genuinely impaired individuals have also been used to establish the clinical utility of this measure. One such study found that community volunteers and clinical cases making a genuine effort scored between 98 and 98.6 on average when completing the primary subscales (standard deviations ranged from 2.8 to 3.1) whereas simulators participating in various scenarios scored between 62 and 70.9 (standard deviations ranged from 12 to 16.5) on average. Information regarding the sensitivity and specificity of this measure are provided above.

Fake Bad Scale (FBS). The FBS was developed by Paul Lees-Haley (1991) using items from the MMPI-2. It consists of 43 true-false items that largely relate to defensiveness and somatic concerns. The test publishers recommend that malingering be suspected at a cut-score of 22 and that a cut-score of 28 raise very significant concerns regarding the validity of reported symptoms (2008). It is typically only scored within the context of a full MMPI-2 administration, and there is no normative data regarding the reliability of this scale in a shortened form of the MMPI-2; therefore, the entire test was administered. Typical administrations lasted one hour to one hour and fifteen minutes. The university students met the basic requirements for completing this test as only a sixth grade reading level is required (Graham, 2006).

The original validation of the FBS was performed with a sample of the author's clients – all of which were participants in litigation after sustaining presumed cognitive impairment (Lees-Haley, et al., 1991). The author boasted an accuracy rate of 93 percent when using a cut-score 20. However, this level of accuracy has not been replicated in any of the follow-up studies on this measure. This may be due, in part, to the author's original selection criteria for malingering which was presumably done independently of the FBS scores. The author vaguely referred to a selection process that involved identifying clients that "appeared clearly to be malingering" (p. 205). In a recent study in which 26 malingerers were compared to 29 head injured individuals, the FBS displayed a sensitivity of .808 and specificity of .862 when a cut-score of 22 was used (Larrabee, 2003). When a cut-score of 28 was used, the FBS displayed a sensitivity of .5 and a specificity of .966. However, the groups used in this study were subsamples from a clinical practice leaving open the question of whether the selection criteria for each group may have led to a systematic bias. Also, questions regarding the widespread application of this measure have been raised due to its tendency to produce high numbers of false positives in some settings (Butcher, et al., 2008).

Procedures. Participants in the malingering condition were given a set of instructions that directed them to simulate difficulty in learning while completing the measures. Participants in the control condition were instructed to complete the measures to the best of their abilities (see Appendix). Participants scheduled appointments to complete all measures at a single appointment. The MMPI-2 was administered during the waiting periods between immediate recognition trials and retention trials on the WMT and TOMM to prevent participants from engaging in confounding interference tasks.

The instructions provided to subjects in the malingering condition were similar to those employed in previous studies (Elhai, et al., 2007; Haines & Norris, 2001). These instructions were intended to provide participants with a scenario so that they could approximate real life conditions. Malingering was simulated by providing significant monetary incentives to participants that “trick” the evaluator by successfully simulating cognitive impairment. Participants in the malingering condition were informed that they should simulate the symptoms of a learning disability or ADHD without being “too obvious” in their attempt to exaggerate impairment so as to evade detection. Participants in the control condition were provided incentive to put forth a good faith effort and to respond candidly on the MMPI-2 by entering them in a raffle on the condition that their results did not indicate random responding.

One study suggested that while monetary incentives in analogue studies are much smaller than the incentives to malinger in real life situations, they seem to provide adequate motivation to change responding behavior (Elhai, et al., 2007). Because there are no set guidelines for what constitutes "enough" incentive, a small pilot study was conducted in which 25 students in an undergraduate psychology course were surveyed. The survey indicated that the students would expect an average of $17 dollars to complete the study, and that the prize amounts offered ($75, $100, and $150 for each condition) would provide adequate motivation for both simulators and controls. Thus, each participant was given a $15 gift card from the university bookstore and a coupon for a free smoothie at a local shop in addition to the promise of having a chance to win one of the larger prizes if they carefully followed the instruction set.

Manipulation checks. In order to assess internal validity, three manipulation checks were applied. First, results from all three measures were analyzed with a MANOVA to verify that the instruction set had a statistically significant effect. The mean scores of the malingering group differed significantly ($p < .001$ for all measures) from the mean scores of the control group.

The second manipulation check took the form of a brief questionnaire administered at the conclusion of the testing. Participants in the malingering condition were asked to rate their agreement with three statements on a 5 point scale. These statements assessed their understanding of the training materials, their belief that they had accurately portrayed a learning problem, and whether they were motivated by the prize money to respond as instructed. Their ratings averaged 4.3, 4.1, and 3.8 respectively indicating their general agreement with the statements. Participants in the control condition were also asked to rate their agreement with three statements on a 5-point scale. These statements assessed their understanding of the training materials, whether they put forth a good faith effort, and whether they were motivated by the prize money to respond as instructed. Their ratings averaged 4.9, 4.9, and 4.0 respectively indicating strong to moderate agreement with the statements

A third manipulation check was applied to the control group only using the VRIN and TRIN subscales on the MMPI-2 to assess for random responding, indiscriminately marking false, or indiscriminately marking true. Only one subject was excluded due to an elevation on the TRIN (t-score of 79) which corresponded with a tendency towards indiscriminately marking true. This manipulation check was not applied to the malingering condition as participants reported using these strategies to simulate academic difficulty.

Results

Once all of the subjects were tested, several indices of psychometric efficiency were calculated for each measure using their published cut-scores, and are presented in Table 1. More specifically, the efficiencies were calculated for trial two of the TOMM, the primary subscales of the WMT, and for the FBS. Two cut-scores were used for the FBS, one at a score of 22 and a more conservative cut-score of 28.

Base rates have long been acknowledged to have a substantial impact on the positive predictive value and the negative predictive value of a given test (Meehl & Rosen, 1955). Because the base rates of malingering will likely vary from university to university, the positive predictive values and the negative predictive values were projected for each measure at varying base rates. This was accomplished using an adaptation of Bayes' Theorem proposed by Glaros and Kline (1998), and the sensitivity and specificity calculated within this study. The data presented in Table 2 is intended to allow for greater generalization.

Table 1

Sensitivity, Specificity, Positive Predictive Values, and Negative Predictive Values at Published Cut-Scores

	TOMM2	WMT IR	WMT DR	FBS 22	FBS 28
Cut-score	45*	82.5**	82.5**	22***	28***
Sens.	0.57	0.90	0.87	0.33	0.00
Spec.	1.00	1.00	1.00	0.97	1.00
PPV	1.00	1.00	1.00	0.91	0.00
NPV	0.69	0.91	0.88	0.58	0.49

Note. * Less than; ** less than or equal; *** greater than or equal

Next, an exploratory analysis was performed using a receiver operating characteristic curve (ROC analysis) to search for possible cut-scores that would improve sensitivity without significantly reducing specificity for any of the measures used in this study. This analysis was particularly relevant to the TOMM as previous studies suggest that its cut-score may be too conservative (Greve, Bianchini, & Doane, 2006), and that trial one of this measure can be an effective screener (Bauer, et al., 2007).

Table 2

Positive predictive values, and negative predictive values adjusted to varying base rates using Bayes' Theorem

	TOMM2		WMT IR		WMT DR		FBS 22		FBS 28	
Base Rate	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
50%	1.00	0.70	1.00	0.91	1.00	0.88	0.91	0.59	0.00	0.50
40%	1.00	0.78	1.00	0.92	1.00	0.92	0.87	0.69	0.00	0.60
30%	1.00	0.84	1.00	0.96	1.00	0.95	0.81	0.77	0.00	0.70
20%	1.00	0.90	1.00	0.98	1.00	0.97	0.71	0.85	0.00	0.80
10%	1.00	0.96	1.00	0.99	1.00	0.99	0.52	0.93	0.00	0.90
5%	1.00	0.98	1.00	0.99	1.00	0.99	0.34	0.97	0.00	0.95

Note. See Glaros and Kline (1998) for an explanation of Bayes' Theorem.

The ROC analysis revealed that if scores for trial one of the TOMM, trial two of the TOMM, Immediate Recall of the WMT, and Delayed Recall of the WMT were considered failure at 44, 49, 92.5, and 95 respectively, it would result in an increase in sensitivity without any decline in specificity. Because an emphasis is typically placed on avoiding false positives (suggesting someone is malingering when they are not) in this context, an "optimal" cut-score did not emerge for the FBS because sensitivity was drastically reduced as specificity reached acceptable levels.

Additionally, ROC analysis creates a curve by plotting sensitivity on the Y axis and 1 – specificity on X axis. The area under this curve is an index of a measure's efficiency, and possible values range from .5 (equivalent to chance when predicting a dichotomous variable) to 1 (perfect prediction). Of the WMT subtests, Immediate Recall had the highest value for area under the curve (.985) indicating that it was able to distinguish between simulators and controls with near perfect accuracy. Of the TOMM subtests, Trial 1 had the highest value for area under the curve (.957) indicating that it was also able to distinguish between simulators and controls with very high accuracy. Finally, the FBS had the least amount of area under the curve (.760 when using raw scores, .775 when using t-scores) of the three measures examined.

A point-biserial correlation was employed to determine the degree to which scores on the FBS were related to group membership (e.g. simulator or control). A preliminary analysis was performed to ensure that the distribution of FBS scores did not violate assumptions of normality. There was only a medium, positive correlation between these two variables, $r_{pb} = .45$, $n = 59$, $p < .001$, with high scores on the FBS being associated with a stronger likelihood of simulating malingering.

Finally, a logistic regression was performed to examine whether prediction is improved by using more than one of the measures of malingering studied in a test battery. Prior to executing the logistic regression, correlations between the subtests and the blocks were examined using Spearman's Rho. The correlation matrix presented in Table 3 indicated that while all the measures were predictive of malingering, attempting to include all of them in the model while performing logistic regression would result in problematic results due to mulicollinearity among the TOMM and WMT indices. Thus, only immediate recall of the WMT Immediate Recall (emerged as the best predictor in the ROC analysis) and the FBS were considered for model.

Table 3

Spearman's Rho between Subtests and Block

	TOMM1	TOMM2	WMT IR	WMT DR	FBS
Block	-.0800**	-0.779**	-0.858**	-0.842**	0.452**
TOMM1		0.863**	0.705**	0.846**	-0.314*
TOMM2			0.772**	0.848**	-0.332*
WMT IR				0.880**	-0.405*
WMT DR					-0.428*

Note. Simulated impairment coded as 1 for "Block"; *Significant at 0.05 level (two-tailed); ** significant at the 0.01 level (two-tailed)

The model generated during the logistic regression was highly significant, χ^2 (2, N = 59) = 66.75, p < .0005, indicating that the model was able to distinguish between simulators and controls. The model explained between 67% (Cox and SnellR^2) and 90% (NagelkerkeR^2) of the variance in block membership, and correctly classified 96.6% of the participants. The only variable entered into the model was the immediate recall subtest of the WMT (Wald = 4.98, df = 1, p = .026) because FBS did not significantly improve prediction.

Study 2

Method

Participants. Data was collected from clients (n = 121) during the course of psycho-educational evaluations requested after self-referral to the university accessibility center. Clients presented with a variety of academic difficulties such as problems with math (34.7%), reading

(61.2%), writing (44.6%), memory (5.8%), attention (46.3%) and/or general academic struggles (5%). Additionally, 62% of the clients presented with two or more problem areas, 26% of the clients presented with three or more problem areas. Clients were routinely informed that they were being evaluated for a perceived academic disability and to determine the appropriateness of accommodative services. Their genders were 54% male and 46% female. Data from one client was excluded from the analysis due to a sleep disorder having invalidated his results.

Measures. Scores from the MMPI-2 (n = 94), TOMM (n = 116), and WMT (n = 104) were used for the data analysis. Sample sizes differ due to missing data; on occasion test administrations were begun, but not finished, due to logistic considerations such as scheduling. If an incomplete administration yielded a valid score for a given subtest, it was included in the final analysis.

Procedures. The clients were university students enrolled at least part-time at a large, private university. Test administrations were scheduled for no longer than two hour periods. Malingering assessment was typically administered at the beginning of the testing periods, and most students attended two to three testing periods.

Results

Possible malingering was operationalized as a score that fell below (or above in the case of the FBS) either a published cut-score or an optimized cut-score generated in study 1. The results of this operationalization are presented in Table 4.

Table 4

Failure Rates by Measure within a Clinical Sample

	TOMM1	TOMM2	WMT IR	WMT DR	FBS 22	FBS 28
n	116	114	104	101	94	94
Published Cut-Scores						
Failure #	--	2	10	8	29	9
Failure %	--	1.8	9.6	7.7	31	9.6
Optimized Cut-Scores						
Failure #	13	3	26	29	--	--
Failure %	11	2.6	25	29	--	--

Note. “Failures P” = total using published cut-scores; “Failures A” = total using optimal cut-scores; -- Value not given if there is no published or optimal cut-score available.

Given that the WMT IR provided the most robust statistics in study 1, it was used to draw comparisons with the other measures. Interestingly, the FBS was the only other subtest that predicted a similar rate of failures as the WMT IR. In order to assess whether the FBS was assessing the same construct; a Pearson’s chi-square was computed for both cut-scores and presented in tables 5 and 6. The results of the analysis were χ^2 (1, N =85) = 1.084, p = .35 when the cut score was placed at 22 and χ^2 (1, N =85) = .631, p = .43 when the cut-score was placed at 28, and therefore non-significant.

Table 5

Cross-tabulation for Failures on WMT Immediate Recall and FBS 22

	FBS 22		
WMT IR	Pass	Failure	Total
Pass	45	17	62
Failure	14	9	23
Total	59	26	85

Table 6

Cross-tabulation for Failures on WMT Immediate Recall and FBS 28

	FBS 28		
WMT IR	Pass	Failure	Total
Pass	56	6	62
Failure	22	1	23
Total	78	7	85

Discussion

The data seem to support the hypothesis that the TOMM would display excellent positive predictive power at the published cut-scores. Thus, the TOMM also displayed excellent specificity. Many participants reported in a post-test questionnaire that the TOMM was extremely easy, and that it was therefore difficult to miss items without feeling that their attempts at simulation were obvious to the administrator (a standardized administration of this measure requires the evaluator to be present at all times and to give direct feedback on the subjects performance). Unfortunately, this appears to have led to relatively modest sensitivity among both the participants in the analogue design and within the clinical sample. However, the findings of these studies also suggest that the sensitively of the TOMM could be improved by creating a cut-score of 45 for trial 1, and that trial 1 could be used as an effective screener for malingering. This would be a welcome tool for many clinicians as it relatively quick and simple to administer.

The hypothesis that the WMT would be more sensitive than the TOMM while displaying similar sensitivity was also supported. The data from the simulation design indicated that the WMT was much more sensitive than the TOMM at the published cut-scores for each test, and that WMT did not produce any false positives. This is particularly important in this context as administrators are typically loath to deny accommodative services from those who qualify. The improved sensitivity of the WMT over the TOMM may be attributable to its format. Students reported in a post-test questionnaire that it was easier to feign impairment while completing this measure because it was mostly computer administered (i.e. no direct supervision by an administrator), and that it appear to be slightly more difficult than the TOMM.

The third hypothesis that the FBS would be poorly correlated with malingering due to problems with construct validity and that it would display poor positive predictive value was only partially supported. The point-biserial correlation of .45 between FBS t-scores and subject block (simulator or control) suggests that the FBS is in fact moderately correlated with malingering. However, as Table 2 illustrates, the positive predictive value of the FBS drops dramatically as the base rate falls to levels expected levels in the context. Moreover, when the more conservative cut-score of 28 was used, this subscale did not improve upon chance alone.

As noted in Table 4, the failure rate on the FBS ranges from 9.6 - 31% depending on the cut-score used. This is somewhat surprising given that none of the participants instructed to simulate impairment scored above 28 on the FBS. Further, when failures on the FBS and WMT were compared within the clinical sample, the results of the Pearson's chi-square suggested that the measures were not assessing the same construct. In isolation, this analysis would not rule out the possibility that the FBS was merely predicting a unique portion of the variance in scores due to malingering which in turn resulted in identifying a unique set of university students. However, given that the FBS did not improve prediction during study 1 when the logistic regression was performed, it seems unlikely that the individuals scoring high on the FBS were merely missed by the WMT. Rather it seems more likely that the FBS scores among the clinical sample are conflated by a tertiary variable such as somatic complaints or general maladjustment (Butcher, Arbisi, Atlis, & McNulty, 2003).

Finally, the hypothesis that the base rate of malingering would be placed at approximately 10% may have been unsupported. Failure rates on the various measures and subtest examine range from 1.8% to 31%. Given that WMT displayed the strongest psychometric properties, its failure rate seems to be the best estimate of the base rate of feigned impairment. The failure rate for the immediate recall subtest at the published cut-score was 10% and the failure rate using the optimal cut-score generated from the data collected in study 1 places the estimate at about 25%. While this is rather large range, it does suggest that the base rate of malingering much lower when compared to results obtained from subjects in forensic neuropsychological studies. One possible explanation for the lowered based rate is that students seeking accommodations for disability do not have as great of a potential gain as do clients participating in litigation for financial awards.

Although it appears that a significant percentage of the students presenting at the accessibility center studied are feigning impairment, it should be noted that this does not necessarily indicate their overall impairment was so distorted as to invalidate their claim for accommodations. Additionally, students seeking evaluation and accommodation for learning disabilities can largely be considered to provide an accurate accounting of their condition via their evaluations and thus should be awarded accommodations if test results support them. Nevertheless, the findings of this study suggest that evaluation results should be examined carefully and that the awarding of accommodations should be contingent upon valid testing. Further study is needed to validate these measures in other universities and to further establish the base rate of malingering in this context.

Limitations

The inability to approximate avoidance of negative consequences as incentives has been labeled one of the "most troublesome" threats to validity in analogue designs (Rogers & Cruise, 1998). Many students presenting for psycho-educational evaluations are likely to be seeking help in order to avoid failure rather than to obtain the benefit of accommodations. The inability to incentivize students with the threat of punishment (failing courses, not getting into graduate school, etc…) may result in differing specificity and sensitivity of measures evaluated in this study between the analogue group and the clinical group. It is hoped that this threat will be minimized by efforts to approximate realistic circumstances through the scenario provided to participants. Another threat to internal validity that relates specifically to the FBS is that there may be group differences in general maladjustment between the analogue group and the clinical group. Given that the FBS is related to maladjustment, group differences in somatic complaints are likely to have resulted in elevated failure rates within the clinical group making it difficult to draw direct comparisons. Our subjects were drawn from a conservative religious university and thus may have had moral and societal reservations regarding any form of deception or poor effort. This argues for the need for follow-up studies with a more heterogeneous subject pool to further establish the base rate of malingering within university settings. Finally, due to inadequate sample size, this study will not attempt to address gender differences in test accuracy.

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*. (4th ed., text revision). Arlington.

Arbisi, P. A., & Butcher, J. N. (2004). Failure of the FBS to predict malingering of somatic symptoms: Response to critiques by Greve and Bianchini and Lees Haley and Fox. *Archives of Clinical Neuropsychology, 19*, 341-345.

Armstrong, D. (2008). Personality check: Malingerer test roils personal-injury law. *Wall Street Journal* A. 1.

Bauer, L., O'Bryant, S. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (2007). Examining the Test of Memory Malingering Trial 1 and Word Memory Test Immediate Recognition as screening tools for insufficient effort. *Assessment, 14*(3), 215-222.

Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 27-49). New York: Guilford Press.

Brennan, A. M., & Gouvier, W. D. (2006). Are we honestly studying malingering? A profile and comparison of simulated and suspected malingers. *Applied Neuropsychology, 13*(1), 1-11.

Butcher, J. N., Arbisi, P. A., Atlis, M. M., & McNulty, J. L. (2003). The construct validity of the Lees-Haley Fake Bad Scale: Does this measure somatic malingering and feigned emotional distress? *Archives of Neuropsychology, 18*(5), 473-485.

Butcher, J. N., Gass, C. S., Cumella, E., Kally, Z., & Williams, C. L. (2008). Potential for bias in MMPI-2 assessments using the Fake Bad Scale (FBS). *Psychological Injury and Law, 1*(3), 191-209.

Cohen, J. (1988). *Statistical power analysis in the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J. M., & McCaffrey, R. J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology, 20*(2), 191-198.

Dawes, R. M. (1962). A note on base rates and psychometric efficiency. *Journal of Consulting Psychology, 26*(5), 422-424.

Elhai, J. D., Butcher, J. J., Reeves, A. N., Baugher, S. N., Gray, M. J., Jacobs, G. A., et al. (2007). Varying cautionary instructions, monetary incentives, and comorbid diagnostic training in malingered psychopathology research. *Journal of Personality Assessment, 88*(3), 328-337.

Fox, D. D., Gerson, A., & Lees-Haley, P. R. (1995). Interrelationship of MMPI-2 validity scales in personal injury claims. *Journal of Clinical Psychology, 51*(1), 42-47.

Frazier, T. W., Frazier, A. R., Busch, R. M., Kerwood, M. A., & Demaree, H. A. (2008). Detection of simulated ADHD and Reading Disorder using symptom validity measures. *Archives of Clinical Neuropsychology, 23*, 501-509.

Gavett, B. E., O'Bryant, S. E., Fisher, J. M., & McCaffrey, R. J. (2005). Hit rates of adequate performance based on the Test of Memory Malingering (TOMM) Trial 1. *Applied Neuropsychology, 12*(1), 1-4.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment, 14*, 196-208.

Gervais, R. O., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19*, 475-487.

Glaros, A. G., & Kline. (1998). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology, 44*(6), 1013-1023.

Graham, J. R. (2006). *MMPI-2: Assessing personality and psychopathology* (Fourth Ed.). New York: Oxford University Press.

Green, P. (2005). *Green's Word Memory Test*. Edmonton: Green's Publishing Inc.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology, 9*(3), 189-207.

Green, P., Lees-Haley, P. R., & Allen, L. M. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Neuropsychology, 2*(3), 97-124.

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

Greiffenstein, M. F., Baker, W. J., & Gola, T. (2002). The Fake Bad Scale in atypical and severe closed head injury litigants. *Journal of Clinical Psychology, 58*(12), 1591-1600.

Greiffenstein, M. F., Fox, D. D., & Lees-Haley, P. R. (2007). The MMPI-2 in Detection of noncredible brain injury claims. In K. B. Boone (Ed.), *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective* (pp. 210-235). New York: Guilford Press.

Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis. *Journal of Clinical and Experimental Neuropsychology, 28*, 1176-1190.

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.

Harrison, A. G., Edwards, M. J., & Parker, K. (2007). Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology, 22*, 577-588.

Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing: Neuropsychological malingering and The Word Memory Test. *Archives of Clinical Neuropsychology, 17*, 709-714.

Kazdin, A. E. (2003). *Research design in clinical psychology* (Fourth Ed.). Boston: Allyn and Bacon.

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

Larrabee, G. J. (2003). Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *The Clinical Neuropsychologist, 17*(1), 54-68.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*, 666-679.

Latham, P. H. (2005). Learning disabilities and the law: After high school: An overview for Students. Retrieved February 15, 2009, from http://www.ldanatl.org/aboutld/adults/civil_rights/law.asp

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.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting Scores. *Psychological Bulletin, 52*(3), 194-216.

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology, 29*(3), 308-318.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1094-1102.

Pearson Assessments (2008). MMPI-2; FBS (Symptom Validity) scale bibliography. Retrieved June, 2010, from http://www.pearsonassessments.com/NR/rdonlyres/A25DB8F8-435F-4066-801B-B641978A97DA/0/mmpi2FBS.pdf

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology, 18*, 379-395.

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment, 10*(1), 10-20.

Rogers, R. (1990). Development of a new classificatory model of malingering. *Bulletin of the American Academy of Psychiatry & the Law, 18*(3), 323-333.

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., Harrell, E. H., & Liff, C. D. (1993). Feigning Neuropsychological Impairment: A Critical Review of Methodological and Clinical Considerations. *Clinical Psychology Review, 13*, 255-274.

Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior, 22*(4), 353-367.

Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15*(4), 349-259.

Ruocco, A. C., Swirsky-Sacchetti, T., Chute, D. L., Mandel, S., Platek, S. M., & Zillmer, E. A. (2008). Distinguishing between neuropsychological malingering and exaggerated psychiatric symptoms in a neuropsychological setting. *The Clinical Neuropsychologist, 22*, 547-564.

Simon, M. J. (2007). Performance of mentally retarded forensic patients on the Test of Memory Malingering. *Journal of Clinical Psychology, 63*(4), 339-344.

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

Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology, 23*, 521-553.

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptoms exaggeration by college adults in Attention-Deficit Hyperactivity Disorder and Learning Disorder assessments. *Applied Neuropsychology, 14*(3), 189-207.

Tombaugh, T. N. (1996). *Test of Memory Malingering: TOMM*. New York: Multi-Health Systems Inc.

Weinborn, M., Orr, T., Woods, S. P., Conover, E., & Feix, J. (2003). A validation of the Test of Memory Malingering in a forensic psychiatric setting. *Journal of Clinical and Experimental Neuropsychology, 25*(7), 979-990.

Appendix

Instructions for the Malingering Condition:
Malingering is the fabrication or gross exaggeration of symptoms in order to get something you want. You will be asked to simulate malingering while you participate in this study. You will receive compensation (gift cards amounting to approximately $15) regardless of your performance. However, you will win one of three larger prizes (150, 100, and 75 dollars) if you are among the top three “performers” meaning you have been successful at tricking the examiner into believing you have a genuine problem (1 in 10 chance). Remember do not tell the examiner you have been asked to malinger. The scenario you are to simulate is presented below:

> Over the last two semesters, your grades have been lower than your expectations. You have spoken to your professors about your concerns, but they feel they cannot offer you any more help without a written note from the University Accessibility Center stating that you having been diagnosed with learning disability or ADHD.As a result, you want proof that you have a genuine disability and you are willing to perform below your abilities on any tests you are given during an evaluation to make sure your needs recognized. Remember, you truly believe you should receive more help from your professors, and that they are being unfair by not being more accommodating. Before going in to be assessed, you prepare yourself to exaggerate your difficulties on a series of psychological tests, and discover that the examiner has methods for detecting obvious attempts at distortion. Therefore, when you actually take the tests, you find ways of letting the examiner know you have a disability by performing low enough to let him/her know you have a realistic problem. Also when given a questionnaire about your personality characteristics, you mark items in such a way that they’ll know you have a problem without being too outrageous.

Instructions for the Control Condition:
You will be asked to complete several psychological tests today. Please give your best effort and respond honestly. You will receive compensation (gift cards amounting to approximately $15) regardless of your performance. However, you will have a chance to win (1 in 10) one of three larger prizes (150, 100, and 75 dollars) if you complete the evaluation to the best of your abilities. Completing the evaluation to the best of your abilities means that you have given a good faith effort, and that you responded honestly on all measures. Remember, your profile on the personality measure given today will not be interpreted other than to determine whether you have responded honestly and consistently. Remember, the examiner has means of detecting random responding or poor effort on all the measures.

Feedback Questionnaire (Experimental Group):

1- - - - - - - -2 - - - - - - - - - 3 - - - - - - - - -4- - - - - - - - - -5
Strongly Disagree Neutral Strongly Agree

Using the scale above, please respond to the following items.

1) I understood the training materials. ____
2) I accurately portrayed a learning disorder._____
3) I was motivated to be an effective malingerer by the prize money offered. ____

Did you have any familiarity with the measures you completed today? If so, do you believe it had any impact on your performance? Please explain.

Please think carefully about the strategies you <u>actually used</u> to malinger on the measures you have been administered today. Based on what you actually did, rate each of the times below.

	I did not use this strategy				A major strategy I used a lot
1. I answered more slowly or hesitated when answering	1	2	3	4	5
2. I intentionally chose wrong or unusual answers	1	2	3	4	5
3. I answered in a somewhat random fashion	1	2	3	4	5
4. I didn't pay much attention to the stimulus materials about which I was to answer questions	1	2	3	4	5
5. I picked a particular psychological problem and tried to imagine how such a person would answer	1	2	3	4	5

6. Did you use any other strategies not listed above? If so please describe.

Feedback Questionnaire (Control Group):

1- - - - - - - -2 - - - - - - - - - 3 - - - - - - - - -4- - - - - - - - - -5
Strongly Disagree Neutral Strongly Agree

Using the scale above, please respond to the following items.

1) I understood the training materials. ____
2) I put forth good effort._____
3) I was motivated to give good effort/respond honestly by the prize money offered. ____

Did you have any familiarity with the measures you completed today? If so, do you believe it had any impact on your performance? Please explain.

Any other comments you wish to share can be written in the space below.