2010

Factors Influencing the Face Validity of Effort Tests: Timing of Warning and Feedback

Haley Gail Trontel
The University of Montana

FACTORS INFLUENCING THE FACE VALIDITY OF EFFORT TESTS: TIMING OF WARNING AND FEEDBACK

By

HALEY GAIL TRONTEL

B.A., University of Puget Sound, Tacoma, WA

Professional Paper

presented in partial fulfillment of the requirements
for the degree of

Master of Arts in Psychology

The University of Montana
Missoula, MT

December 2010

Approved by:

Perry Brown, Associate Provost for Graduate Education
Graduate School

Stuart Hall, Ph.D., Chair
Psychology

Daniel Denis, Ph.D., Co-Chair
Psychology

Annie Sondag, Co-Chair
Education and Human Sciences

Trontel, Haley Psychology

Face Validity of Effort Tests

Chairperson: Stuart Hall, Ph.D.

Co-Chairperson: Daniel Denis, Ph.D.

Co-Chairperson: Annie Sondag, Ph.D.

The assessment of client effort during neuropsychological evaluation is of high importance. Two experiments were designed to assess factors that may influence effort testing during neuropsychological assessment. Participants for both experiments were undergraduate students without a history of neurological conditions, mental health concerns, or current problems with alcohol or drug use. The goal of Experiment 1 was to determine whether the timing of a warning that some tests may detect faking would influence (the face validity of and performance on) effort and standard neuropsychological measures. All participants were administered the Test of Memory Malingering (TOMM) and Memory for Complex Pictures (MCP) as well as a brief battery composed of standard cognitive measures. Following administration of all tests, participants completed a questionnaire assessing their perception of the purpose of each measure. Results from Experiment 1 reveal that Early Warning CBIS endorsed lower face validity for effort measures than the other three groups. Performance on the effort measures was not significantly different across groups.

Experiment 2 examined the role of visual feedback on individuals' performance on effort testing. Results from Experiment 2 revealed that the use of visual feedback does not influence effort test performance.

Table of Contents

Abstract **2**
Introduction **5**
The Effect of External Incentive on Effort **6**
The Effect of Coaching on Effort **7**
Measures Designed to Detect Inadequate Effort **10**
The Effect of Warning on Inadequate Effort **13**
The Effect of Warning and Coaching on Face Validity **14**
Experiment 1 **15**
Purpose of Experiment 1 **15**
Hypotheses for Experiment 1 **16**
Method for Experiment 1 **17**
Participants 18
Materials 18
Medical and Health History Questionnaire 18
Memory for Complex Pictures (MCP) 19
Test of Memory Malingering (TOMM) 20
Standard Neuropsychological Measures 20
Role-Play Termination Instructions (RPT) 21
Manipulation Check Questionnaire (MCQ) 21
Face Validity Questionnaire (FVQ) 21
Procedure 22
Results for Experiment 1 **23**
Power for Experiment 1 23
Demographic Information 24
Questionnaire Responses for Effort Tests 25
Face Validity Questionnaire (FVQ) 25
Performance on Effort Measures 28
Memory for Complex Pictures (MCP) 28
Test of Memory Malingering (TOMM) 28
Questionnaire Responses for Standard Tests 30
Face Validity Questionnaire (FVQ) 30
Performance on Standard Measures 33
WAIS-III Digit Symbol-Coding 33
WAIS-III Digit Span 33
Finger Tapping 33
Trails A 34
Trails B 34
Manipulation Check 35
Experiment 2 **37**
Purpose of Experiment 2 **38**
Hypotheses for Experiment 2 **38**
Method for Experiment 2 **38**
Participants 38

Materials 39
Medical and Health History Questionnaire 39
Memory for Complex Pictures (MCP) 39
Manipulation Check Questionnaire (MCQ) 39
Procedure 40
Results for Experiment 2 **41**
Power for Experiment 2 42
Demographic Information 42
Performance on Neuropsychological Measures 42
Memory for Complex Pictures (MCP) 42
Questionnaire Results of Experiment 2 43
Manipulation Check Questionnaire (MCQ) 43
Discussion **45**
Face Validity of Effort Tests **46**
Performance on Effort Tests **48**
Face Validity of Standard Tests **50**
Performance on Standard Tests **51**
Performance on Effort Tests with and without Feedback **52**
Conclusion **53**
References **55**

Appendix A **61**
Appendix B **62**
Appendix C **63**
Appendix D **64**
Appendix E **65**
Appendix F **66**

Factors Influencing the Face Validity of Effort Tests

Obtaining optimal effort during neuropsychological evaluations is a primary goal. Accurate assessments require that individuals are performing to the best of their ability at all times. In some circumstances, individuals do not perform their best, challenging the accuracy of the neuropsychological assessment. There are a variety of reasons a person may not put forth their best effort during neuropsychological assessment.

A large area of research on inadequate effort on neuropsychological tests is focused on malingering. Malingering is the intentional production of false or greatly exaggerated symptoms for the purpose of attaining some external reward (American Psychiatric Association, 1994). Resnick (1997) further classified malingering into three types, “pure malingering,” partial malingering,” and “false imputation,” or feigning disability. Examinees may also be simply unmotivated to perform to the best of their ability, regardless of an external incentive.

The terminology in the literature of effort during neuropsychological testing is quite varied. The terms nonoptimal effort, suboptimal effort, poor effort, biased responding, dissimulation, and negative response bias are all used to describe inadequate effort during assessment (e.g. Bernard, 1990; Gorny & Merten, 2005, Iverson, 2006, Kerr, et al., 1990). The term exaggeration, or symptom exaggeration, is often used to describe inaccurate responses on psychological assessments. Iverson (2006) recommends the use of *poor effort* to describe underperforming on neuropsychological tests. Therefore, responses that do not indicate optimal performance will be termed poor or inadequate effort in this paper, regardless of the reason.

Poor effort is a significant problem in neuropsychology. Youngjohn, Burrows, and Erdal (1995) speculate that about half of all workers' compensation claims involve false cognitive deficits. A survey of American Board of Clinical Neuropsychology members found high base rates, from 19% to 30% of malingering and symptom exaggeration in cases involving personal injury, disability, and criminal referrals (Mittenberg, Patton, Canyock, & Condit, 2002). In other research, persons with injury claims alleging mild head injury, fibromyalgia, chronic pain, neurotoxicicity, and electrical injury claims had base rates of malingering from 22% to 39% (Mittenberg, Patton, Canyock, & Condit, 2002). Rates of inadequate effort are particularly high in forensic settings. Ardolf, Denney, and Houston (2007) reviewed the published literature on inadequate effort on neuropsychological assessment in a criminal forensic setting and found base rates of malingering at 53.4%. Similar results were found by Delain, Stafford, and Ben-Porath (2003), who found 45.3% of their criminal court forensic sample to be malingering on the TOMM as part of the neuropsychological assessment.

Although inadequate effort is relatively common, individuals are unlikely to admit when they do not provide optimal effort on neuropsychological tests. In addition, individuals who are likely to provide inadequate effort do not have distinctive demographic characteristics (Delain, Stafford, & Ben-Porath, 2003). Therefore, the development of measures to detect poor effort as well as an understanding about variables that influence effort are paramount goals in neuropsychological research.

The Effect of External Incentive on Effort

Motivation has an important influence on neuropsychological test performance. In particular, monetary compensation seems to be an important motivating factor. Clients involved in litigation have been found to have greater neuropsychological complaint rates compared to non-litigating clients (Green & Iverson, 2001; Green, Iverson, & Allen, 1999). On neuropsychological tests, such patients may perform substantially below their actual cognitive capacities. Indeed, individuals seeking financial compensation were four times as likely to demonstrate invalid performance on a popular test of effort, the Test of Memory Malingering (TOMM; Moore & Donders, 2004). Flaro, Green, and Robertson (2007) compared failure rates on another common effort test, the Word Memory Test (WMT), among participants who may be motivated to underperform, mild Traumatic Brain Injury (TBI) patients under litigation, with individuals who may be motivated to perform their best, parents undergoing a parenting assessment to seek custody of their children. It was expected that involvement in litigation (seeking monetary compensation) would produce inadequate effort whereas seeking custody would produce good effort. Indeed, the failure rate in the mild TBI sample was 23 times higher than in the group of parents seeking custody. These studies illustrate the importance of incentive and motivation on test performance.

The Effect of Coaching on Effort

Coaching is providing clients with specific information about brain injury symptoms (e.g. memory problems) or how a person with brain injury might perform on tests (e.g. slow reaction time) in such a way that accurate assessment of their performance on neuropsychological measures is hindered (DenBoer & Hall, 2007; Gorny & Merten, 2005; Rose, Hall, Szalda-Petree, & Bach, 1998). Results of a recent survey on member of the National Academy of Neuropsychology (NAN) and the Association of Trial Lawyers indicated that 75% of attorneys said they spend an average of 25-60 minutes coaching their clients about psychological testing and the testing process. The same survey found that 44% of the attorneys wanted to know what specific neuropsychological measures would be used by the psychologist (Wetter & Corrigan, 1995). This might suggest that attorneys are coaching directly to the test.

Coaching can take many forms. Coaching in neuropsychological assessment involves providing information about brain injury to clients or research participants. Individuals may coach themselves by researching information about brain injury without the aid of an attorney. Utilizing information concerning common brain injury symptoms, individuals appear to be able to simulate the performance of patients with authentic neurological impairment (e.g., Rogers, Gillis, Bagby, & Monterio, 1991).

Relatively minimal coaching has been found to be effective in producing sophisticated malingering. For example, Kerr et al. (1990) provided simulators with a magazine article regarding the effects of head injury, finding that the performance of participants in this study closely resembled the performance of participants with an authentic brain injury. Additional studies have found that even individuals with little background knowledge can successfully portray cognitive deficits or pain symptoms on neuropsychological testing (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988; Gervais, Green, Allen, & Iverson, 2001; Heaton, Smith, Lehman, & Vogt, 1978). Approaches used to feign impairments include faking memory loss and adjusting their response time. Tan, Slick, Strauss, and Hultsch (2002) found that coached participants reported feigning memory loss as their main strategy, with slow response time being their second strategy.

The ability for individuals to successfully feign impairment challenges the accuracy of neuropsychological measures, including those designed to detect poor effort. DiCarlo, Gfeller, & Oliveri (2000) found that coaching reduces the likelihood that an individual who is underperforming will be detected by neuropsychological measures. Similarly, Suhr and Gunstead (2000) found that approximately 31% of the uncoached malingerers in their study were detected by a forced-choice measure of malingering, whereas only approximately 6% of the coached malingerers were detected.

Another coaching study done by Rose, et al. (1998) asked participants to pretend they had endured a brain injury while in a car accident. The coached group was also given a description of the effects of head injury and information on how a head injured person may present. Uncoached participants were told only to pretend they had a brain injury. Coached participants performed significantly better than uncoached participants, but still performed worse than controls. Additionally, Rose, et al. (1998) administered a computerized version of the Portland Digit Recognition Test (PDRT-C) to the coached and uncoached malingerers as well as the controls. They found that coached malingerers were better able to avoid detection by the PDRT-C, providing evidence that coaching may increase the possibility that malingerers can evade detection on neuropsychological tests of malingering.

Recently, Gorny and Merten (2005) reported that giving explicit information on malingering measurement was the most effective form of coaching. However, even giving individuals symptom information or a warning against exaggerating raised their scores on measures of effort, although not significantly. Research has also found that clinicians are unable to detect inadequate effort at better than chance levels using their clinical judgment alone (Faust, Hart, Guilmette, & Arkes, 1988). Given the evidence that individuals have motivation to underperform, the ability for individuals to successfully appear brain injured, and that clinician judgment alone cannot accurately detect faking, it is pertinent that measures are developed to accurately detect inadequate effort.

Measures Designed to Detect Inadequate Effort

In one of the first studies to demonstrate that individuals were able to successfully feign impairment, Heaton, Smith, Lehman and Vogt (1978) asked neuropsychologists to determine whether individuals' performances were produced by a malingerer or a real head-injury patient. Neuropsychologists were able to detect malingerers at about chance-level to slightly above chance. Clinician's inability to accurately distinguish inaccurate performance from real performance led to an interest in empirical strategies to detect inadequate effort. Attempts to identify inadequate effort using tests have been around for decades, with Rey's 15-item test being one of the first (Rey, 1964). Researchers tried to use neuropsychological measures, like the Wechsler Memory Scale-Revised, to determine effort (Bernard, 1990). Results from these studies were generally inadequate. However, results from these studies found that the most sensitive tests were those which utilized some type of recognition task (Tombaugh, 1996). Discriminant function based on neuropsychological tests were significantly better at accurately detecting the malingerers, demonstrating the value of objective measures in detecting inadequate effort (Heaton & Rose, 1998).

A large portion of research has been devoted to developing and improving tests specifically designed to detect test performance suggestive of inadequate effort. To uphold the integrity of the results, the measures are intended to look like tests of cognitive abilities (e.g. memory). Symptom Validity Tests (SVT) were designed to identify performance that is significantly poorer than what would be expected even among brain-damaged populations. In other words, clinical populations and populations with rather severe brain injury perform very well on these tests. Therefore, poor performance on these tests by an individual is thought to be indicative of inadequate effort. Hiscock and Hiscock (1989) adapted the SVT model by asking individuals to remember a five-digit number across three different retention intervals. A level of performance that fell significantly below chance was indicative of malingering. The Portland Digit Recognition Test (PDRT; Binder, 1990) uses a similar procedure but the person is asked to count backwards during the retention intervals and different intervals are used. However, some studies have found that both the Hiscock and the PDRT lack adequate sensitivity (Binder & Willis, 1991; Guilmette, Hart, & Guiliano, 1993).

One of the most heavily researched and used symptom validity tests is the Test of Memory Malingering (TOMM, Sharland & Gfeller, 2007; Slick, Tan, Strauss, & Hultsch, 2004; Tombaugh, 1996). The TOMM (Tombaugh, 1996) consists of two learning trials and a retention trial. In each learning trial the examinee is shown 50 line-drawings for 3 seconds each, with a 1-second interval between pictures. The examinee is then shown 50 two-choice recognition panels. These panels consist of a previously shown picture and a distracter picture and the examinee must choose the correct picture. There is a 50% chance of choosing either the previously shown picture or the distracter picture. Below-chance performance may signify inadequate performance. Indeed, individuals who are performing to the best of their ability tend to receive near perfect scores on tests of effort (Tombaugh, 1996; 1997).

Rees, Tombaugh, Gansler, and Moczynski (1998) conducted five validation experiments on the TOMM and found it to be sufficiently sensitive and specific as a test of memory malingering. Similar results were found by Delain, Stafford and Ben-Porath (2003). However, Tan, Slick, Strauss, and Hultsch (2002) concluded that among the TOMM, Victoria Symptom Validity Test (VSVT), and the WMT, the TOMM performed the poorest with respect to accurate classification.

In an effort to advance existing symptom validity measures, a computerized malingering measure, Memory for Complex Pictures (MCP; DenBoer & Hall, 2005; DenBoer, 2007) was recently developed with hopes of improved face validity, specificity, and sensitivity. The MCP's use of digit photographs of complex visual scenes was based on the relatively high and robust visual memory capacity of human beings. Complex visual scenes are more easily remembered than simple visual scenes, increasing the sensitivity of the measure (Nelson, Metzler, & Reed, 1974). The MCP was validated among neurological patients, controls, and simulated malingerers (DenBoer, 2007). The results of the study demonstrated that the MCP, when compared to the TOMM, displayed improved ability to detect malingering as well as face validity. Notably, a group of mixed-clinical patients with demonstrable memory deficits did very well on the MCP, further supporting its potential worth as a useful measure of malingering-detection. In addition to displaying very good psychometric characteristics, the MCP also demonstrated increased efficiency of administration. The current study will further assess the ability of the test to detect malingering and the face validity of this measure as it compares to the TOMM as well as explore specific factors that may influence the face validity of these measures.

The Effect of Warning on Inadequate Effort

Additional research has found that warning participants prior to testing about the presence of measures designed to detect inadequate effort can challenge the accuracy of assessments. A survey of 24 expert neuropsychologists found that almost 38% always provide a warning that some tests may be designed to detect inadequate effort prior to testing. Another 54% of respondents indicated never providing such a warning prior to testing (Slick, Tan, Strauss, & Hultsch, 2004). Another study found that, out of 188 surveyed neuropsychologists, 52% rarely or never provide a warning while 27% often or always provide a warning (Sharland & Gfeller, 2006). The inconsistency in clinician's decisions about when they provide a warning neglects the possible influence of the timing of the warning on examinee test performance.

Johnson and Lesniak-Karpiak (1997) found that malingering simulators who were given a warning were able to avoid detection by improving their performance on malingering measures compared to simulators who weren't warned. Indeed, 45% of the warned malingerers were misclassified as controls. Youngjohn, Lees-Haley, and Binder (1999) found that warning persons of detection methods on malingering procedures did not make them appear exactly like controls; instead, warning may produce more sophisticated test takers. In other words, providing a warning may allow test takers to know the true nature of the test and alter their performance accordingly. Indeed, Youngjohn, Lees-Haley, and Binder (1999) suggest that providing a warning may actually allow individuals to feign deficits more accurately. Providing a warning prior to or after testing appears to be common practice among neuropsychologists. It is therefore imperative to understand how timing of that warning affects not only the face validity, but also the accuracy of the performance on the measures.

The Effects of Warning and Coaching on Face Validity

Face validity is what a test appears to measure to the examinee (Anastasi, 1988). A key challenge in research on inadequate effort is avoiding the possibility that individuals are able to determine that a measure is a test of effort. Bornstein, Rossner, Hill, and Stepanian (1994) examined the relationship between face validity and whether or not the participants were likely to fake their performance on a measure of dependency. Results indicated that the higher the face validity of a cognitive measure, the less easy it was to feign responses on it. In other words, participants who recognized the test as measuring what it was actually intended to measure were less likely to fake their performance. This indicates that high face validity on measures of effort is important to reduce the likelihood of faking.

In a study by Tombaugh (1997), participants felt confident that the TOMM was a measure of memory, indicating high face validity for the measure. Rees, Tombaugh, Gansler, and Moczynski (1998) also found that participants viewed the TOMM as a memory test when it was administered in a battery of neuropsychological tests.

Tan, Slick, Strauss, and Hultsch (2002) compared the face validity of the TOMM, Word Memory Test (WMT), and the Victoria Symptom Validity Test (VSVT). The authors found that the WMT displayed the best face validity, with approximately thirty-one percent of the participants reporting the WMT as a cognitive measure. The TOMM displayed poorer face validity than the WMT and the VSVT. In addition, approximately one-third of participants correctly identified all three instruments as measures of effort. This finding reinforces a need for superior tests to detect inadequate effort, particularly in regards to improved face validity.

Given the influence of coaching, external incentive, and the presence of a warning on neuropsychological test performance, it is crucial to have well-validated effort measures in order to more accurately detect inadequate effort and improve the validity of assessment. In order to accurately test effort, it is important that the measures have high face validity. Both coaching and the presence of warning have been shown to challenge the face validity and the overall integrity of neuropsychological assessments. Although many tests have been designed to detect inadequate effort, new tests with improved sensitivity, specificity, and face validity are necessary. The current study compares the face validity of the TOMM with the Memory for Complex Pictures (MCP; DenBoer, 2007) and examines how timing of warning and the presence of coaching affect performance on these measures.

Purpose of Experiment 1

Experiment 1 has several goals. The first goal, and primary focus, was designed to examine whether or not a warning that the test battery contained tests designed to detect if someone is faking brain damage, prior to administration of the neuropsychological tests versus receiving a warning after administration of the tests, would affect the face validity and scores on the MCP and TOMM as well as the other standard cognitive measures. One half of the participants received the warning prior to administration of the measures and the other half received the warning following administration of the measures.

An additional goal was to compare the ability of the MCP and the TOMM to detect students who are simulating a brain injury during neuropsychological testing (i.e., sensitivity). The two measures were compared in participants' ability to recognize them as tests of effort (i.e., face validity). The final goal was designed to examine whether certain factors (being told to simulate brain injury, being told to perform to the best of their ability) would affect the face validity and scores of the MCP and TOMM.

Hypotheses for Experiment 1

1) It was hypothesized that the MCP will demonstrate higher overall face validity as an actual measure of visual memory than the TOMM regardless of when the warning is delivered. Specifically, on Face Validity Questionnaire (FVQ; Appendix D) it was predicted that the MCP would be endorsed by more participants as an actual measure of neuropsychological functioning than the TOMM.
2) It was hypothesized that participants in the Control groups would obtain significantly higher scores on the MCP and TOMM than the Coached Brain Injury Simulator early warning group (CBIS-early) and Coached Brain Injury Simulator late warning group (CBIS-late).
3) It was hypothesized that participants in the CBIS-early group would obtain significantly higher scores on the MCP and TOMM than the CBIS-late group.
4) It was hypothesized that the MCP and TOMM would demonstrate lower face validity in the CBIS-Early group than in the CBIS-late group. Specifically, the MCP and TOMM would be endorsed as effort tests more frequently on the FVQ in the CBIS-early group than in the CBIS-late group.

5) It was hypothesized that the MCP and TOMM would demonstrate lower face validity in the CBIS-late group than in the control groups. Specifically, the MCP and TOMM would be endorsed as effort tests more frequently on the FVQ in the CBIS-late group than in the control groups.
6) It was hypothesized that participants in the control groups would obtain significantly higher scores on the standard tests (described below) than the CBIS-early group and CBIS-late group.
7) It was hypothesized that participants in the CBIS-early group would obtain significantly higher scores on the standard tests than the CBIS-late group.
8) It was hypothesized that the standard measures would demonstrate lower face validity in the CBIS-early group than in the CBIS-late group. Specifically, the standard measures would be endorsed as efforts tests more frequently on the FVQ in the CBIS-early group than in the CBIS-late group.
9) It was hypothesized that the standard measures would demonstrate lower face validity in the CBIS-late group than in the control groups. Specifically, the standard measures would be endorsed as effort tests more frequently on the FVQ in the CBIS-late group than in the control groups.

Method for Experiment 1

Participants

Sixty University of Montana undergraduate psychology students were recruited to participate in this study. Students were randomly assigned to one of four groups. First, students were assigned to either a Control group or a Coached Brain Injury Simulator (CBIS) group. The Control groups were told to perform to the best of their ability and the

CBIS groups were asked to pretend to perform as though they were brain injured. The students in each of these groups (Control and CBIS) were randomly assigned into either an Early or Late Warning group. The Early Warning group was told that some of the measures may test for faking of brain injury before administration of the neuropsychological measures and the Late Warning group was told after administration of the measures. Students were 18 years of age or older and were excluded from the study if they reported a history of psychological or neurological problems or treatment. The 60 students were given 6 credits toward their respective courses for participation. Each student completed the MCP and the TOMM, as well as additional standard neuropsychological measures.

Materials

Medical and Health History Questionnaire. A medical and health history questionnaire was used to assess participants' history of psychological, neurological, and substance abuse problems (See Appendix A). This measure was used for screening. Specifically, if participants endorsed any neurological or psychological problem they were excluded from the study. Additionally, if participants endorsed three substance abuse questions they were excluded from the study.

Memory for Complex Pictures (MCP). The MCP is a forced-choice, two-alternative measures consisting of complex digital photographs for assessing effort during neuropsychological evaluation. The test begins with a sample trial using 3 digital photographs immediately followed by a recognition trial that pairs the target stimulus pictures with similar foils. Participants are asked to recognize all three sample items correctly before continuing to Trial 1. Participants have three opportunities to correctly complete the sample trial. If they fail all three trials the test is cancelled.

Fifty photographs are presented over the course of two learning trials. During both learning trials the individual is exposed to all 50 photographs presented for 3 sec each with a 1 sec inter-stimulus interval. Immediately following each learning trial is a recognition trial. All pairs are the same in Trails 1 and 2 but ar5e presented in a different order over the course of the two learning trials.

After completing the stimulus presentation trials, participants view a brief screen that provides them with instructions for the recognition trial. The recognition trial immediately follows each presentation trial. In the recognition trial the target stimulus is paired in vertical fashion with a foil and the individual is asked to choose the image that they remember seeing previously. The examinee chooses an image by pressing the “2” key for the top picture of the pair or the “8” key for the bottom picture. As an alternative, the examinee is also allowed to use the keyboard arrows, with ↑ denoting the top picture and ↓ denoting the bottom picture. The foil is another complex visual scene containing similar stimuli to the target scene. The same procedure is followed for the second presentation and recognition trial.

During the recognition trial if the examinee chooses the target stimulus the word “RIGHT” appear in all caps, 18-point, bold font, approximately two inches to the right of the target stimulus. Both the target stimulus and the foil are present for .75 sec. If the examinee selects the foil instead of the target, then the word “WRONG” appears to the right of the foil and both pictures remain on screen for .75 sec. The optional auditory feedback mechanism was not used in this study.   Given that a total of 50 correct responses per trial can be obtained, a total score of 25 or below on Trial 2 represents chance level of performance. A preliminary cutoff score with clinical participants on Trial 2 was found to be 44 (DenBoer, 2007).

Test of Memory Malingering (TOMM). The TOMM is a 50-item, two-alternative, forced-choice measure of client effort used during neuropsychological assessment. The TOMM is a widely used measure, with good validity and reliability (Strauss, Sherman, & Spreen, 2006). Rees, Tombaugh, Gansler, and Moczynski (1998) found that the TOMM demonstrates high specificity and sensitivity and is not affected by demographic variables. This measure consists of two learning trials and a delayed retention trial. Both learning trials contain the same 50 line drawings of common objects, presented in different order in each trial. The trials are administered by the examiner using paper booklets. The learning trial is conducted by presenting each drawing for 3 seconds with a 1 second inter-stimulus interval. Each learning trial is followed immediately by a recognition trial. During the recognition phase, the target stimulus and a foil (another line drawing) are presented in vertical fashion on a small page and the examinee is asked to point to which picture they were shown before. The participants are then provided verbal feedback by saying either "correct" or "incorrect." A score of 25 on Trial 2 represents a chance level of performance and a score of 45 on Trial 2 or the Retention Trial represents cutoff score performance. The Retention Trial will not be included in the current study.

Standard Neuropsychological Measures. Five additional neuropsychological assessment measures were administered along with the MCP and TOMM. The following measures were used: the Digit Symbol-Coding and Digit Span subtests from the

Weschler Adult Intelligence Sacle-Third Edition (WAIS-III; Weschler, 1997), the Trial Making Test—Parts A and B (Reitan & Wolfson, 1993), and the Finger Tapping Test (Halstead, 1947).

Role-Play Termination Instructions (RPT). After completing all standard neuropsychological measures, all participants were given brief written instructions asking them to terminate their role play (if they were engaged in role play) for the remainder of the study (See Appendix B). The RPT allowed participants who were asked to pretend to have sustained brain-injury to answer the subsequent questionnaires honestly.

Manipulation Check Questionnaire (MCQ). After receiving the RPT, participants were given the MCQ (See Appendix C). Designed as a "manipulation check," the MCQ contained three questions designed to make certain all participants included in this study's analysis were able to recall their instructions accurately and indicated that they had put forth at least moderate effort at following their experimental instructions during the study. In order to establish the participants' effort at and success in following their instructions, a Likert-scale ranging from 1 (*didn't try at all)* to 10 (*tried very hard)* was included. Only participants who indicated that they correctly understand their instructions and at least moderately try to follow their instructions were included in the analyses. To determine how successful the participants felt they were in accomplishing their task, a Likert scale ranging from 1 (*not at all successful)* to 10 (*very successful)* was included.

Face Validity Questionnaire (FVQ). After completing the MCQ, participants were given the FVQ (See Appendix D). Participants were asked to indicate what they thought each neuropsychological test was supposed to measure. To determine how confident they were in their response, a Likert-scale ranging from 1 (Not at all confident) to 10 (Extremely confident) was also included. Participants in the late warning group received instructions at the top of the FVQ which indicated that some of tests they took may have been designed to detect faking brain injury (See Appendix E).

Procedure

Participants were brought into the lab and provided one of four types of instructions (CBIS-early, CBIS-late, Control-early, Control-late; See Appendix F). The order of these instructions was randomized. Control participants were asked to perform to the best of their ability. Coached Brain Injury Simulators (CBIS) received the following instructions:

> You are about to take some cognitive tests that examine mental abilities such as attention, memory, thinking, and reasoning skills, and your ability to think quickly.
> While responding to the tests, please pretend that you have experienced brain damage from a car accident involving a head-on collision. You hit your head against the windshield and were knocked out for 15 minutes. Afterwards, you felt "dazed" so you were hospitalized overnight for observation. Because the driver of the other car is at fault, you have decided to go to court to get money from the person responsible. During the next few months following the accident, the negative effects from your head injury disappear. Your lawsuit has not been settled yet, and your lawyer has told you that you may get more money if you look like you are suffering from brain damage.
> As you pretend to be this car accident victim, try to respond to each test as a patient who is trying to appear brain damaged in order to get money from the lawsuit. Thus, your performance on the tests should convince the examiner as well as the people involved in deciding the outcome of your lawsuit that you are still suffering from brain damage. In order to convince these individuals, your brain damage must be believable. *If you succeed in convincing the examiner that you have suffered a brain injury, you will receive two additional experimental credits, for a total of 6 credits.*
> Try to produce the most severe problems that you can *without* making it too obvious to the examiner. Major exaggerations, such as remembering absolutely nothing, are easy to detect. If the examiner does not believe that you have any problems you will not get any money for your head injury. People who have a head injury often have problems paying attention, cannot remember things as well, and do not learn things as easily as

> they did before their injury. They also think a little slower than they used to. Keep this in mind when taking the tests. Remember you are to try to mimic the performance of persons who are truly brain damaged.

Coached Brain Injury Simulators in the early warning group received the same instructions prior to administration of the measures as above. However, the early warning group had the following additional sentence embedded in the instructions, which constitutes the warning:

> It is possible that some of the tests you will take today were designed to detect if someone is faking brain damage.

Participants in the Coached Brain Injury Simulators late warning group received the same additional sentence (warning) after administration of the neuropsychological measures. The warning was imbedded in the Face Validity Questionnaire that appeared after the RPT and MCQ.

All participants were administered the medical and health history questionnaire as a screening measure prior to participation in the study. After receiving the scenario, the neuropsychological measures were administered. The neuropsychological measures, including the MCP and the TOMM were counter-balanced using a latin-square design. These measures were followed by the Manipulation Check Questionnaire and the Face Validity Questionnaire.

Results for Experiment 1

Power for Experiment 1

For the purposes of this analysis power was set at .80 with a significance level of $\alpha = .05$ (Cohen, 1992). Large effect sizes (eta squared) were assumed based on the work of Huskey (2005) and DenBoer (2007), whose results with similar measures yielded R-squared effect size estimates for type of instructions (i.e., group) exceeding .40 and power ranging from .92 to 1.0 (p. 71). Given that mean differences were collected, 15 participants per group were required for the analyses. Provided that a 2 x 2 ANOVA was required to be conducted, a minimum of 60 total participants were needed in Experiment 1. Given the predetermined power, 60 participant were used for this study.

Demographic Information

A total sample size 60 undergraduate students completed the battery of neuropsychological measures. Demographic characteristics are shown in Table 1. No significant differences for gender were found between the four groups, χ^2 (3, 60) = 5.691, *ns*. Group differences for Age and Education were analyzed using two separate one-way ANOVAs. There were significant differences found for Age, *F*(3, 60) = 3.24, *p*<.05. Tukey HSD comparison revealed that participants in the Early Warning Controls were significantly older than Late Warning Controls, *t*(3, 60) = 0.029, *p*<.05. No other significant group differences for age were found. There was no significant difference found for Education, *F*(3, 60) = 0.17, *ns.*

Table 1
Demographic Information for Participants in Experiment 1

	Group				
	Early Warning Controls *(n=15)*	Late Warning Controls *(n=15)*	Early Warning CBIS *(n=15)*	Late Warning CBIS *(n=15)*	χ^2 *or F (df)*
Gender					
Males (n)	5	10	6	4	χ^2(3)=5.69
Females (n)	10	5	9	11	
Age					
M	22.93_a	19.07_b	19.87	19.80	F(3)=3.24*
(SD)	*(6.55)*	*(1.22)*	*(1.85)*	*(2.60)*	
Education					
M	12.40	12.20	12.27	12.33	F(3)=0.17
(SD)	*(1.12)*	*(0.76)*	*(0.59)*	*(0.62)*	

Note. Means in the same row having different subscripts are significantly different at *p*<.05 in the Tukey HSD comparison. * *p<.05*

Questionnaire Responses for Effort Tests

Face Validity Questionnaire (FVQ). Results from the FVQ were examined to determine the face validity of the effort measures. Participants' responses were categorized for each test based on previous research (DenBoer, 2007): (a) Memory, (b) Effort, (c) Learning, (d) Attention/Concentration, (e) Mental Speed, (f) Coordination, (g) Motor Speed, (h) Other, and (i) Don't Know.

The overall face validity of the TOMM was compared with the overall face validity of the MCP regardless of when the warning was presented. A Chi-square was used to compare the number of participants who endorsed the MCP as a test of memory, the number of participants who endorsed the TOMM as a test of memory, and the number of participants who endorsed both the MCP and the TOMM as a test of memory. A Chi-square test for independence indicated a significant difference, χ^2 (2, 60) = 58.62, p<0.001, with significantly more participants endorsing both the TOMM and the MCP as a test of memory than either test individually. The number of participants who endorsed the TOMM as a test of memory versus the number of participants who endorsed the MCP as a test of memory was not significantly different. In other words, most participants viewed both the TOMM and the MCP as tests of memory.

Table 2 presents the frequencies of participant's beliefs regarding the nature of effort tests. Frequencies for the two Early Warning groups represent participant's beliefs regarding the nature of the effort tests when receiving information that some of the tests may be tests of effort *prior* to administration of the tests. Frequencies for the Late Warning groups represent participant's beliefs regarding the nature of the effort tests when receiving information that some of the tests may be tests of effort *after* completing the tests.

A Pearson Chi-Square revealed significant group differences for participants endorsement of the MCP as a test of memory, χ^2 (3, 60) = 8.35, $p < .05$ (*Cramer's V*= .40; medium effect). Posthoc Chi-Square revealed that participants in the Early Warning CBIS group had significantly fewer endorsements of the MCP as a test of memory compared to the other three groups. The three other groups all had large number of participants endorse the MCP as a test of memory, although in all four groups, well over half (57-100%) of the participants endorsed the MCP as a test of memory. No significant differences were found for participants' endorsement of the TOMM as a test of memory.

Significant group differences were not found for participants' endorsement of the MCP as a test of effort as tests cannot be run without any data in three of the four groups. However, only participants in the Early Warning CBIS group indicated that the MCP was a test of effort. Although significant tests cannot be run, this is in line with the hypothesis that individuals in the Early Warning CBIS group would be more likely to see the MCP as a test of effort. However, other explanations for this finding exist and are elaborated on in the discussion. In summary, a very high percentage of participants viewed the effort tests as test of memory.

Table 2

Frequency of Participants Endorsing Cognitive Domains on the FVQ-Effort Tests

	Group				
	Early Warning Controls (n=15)	*Late Warning Controls (n=15)*	Early Warning CBIS (n=15)	Late Warning CBIS (n=15)	χ^2
Memory					
TOMM	11	14	13	14	7.50
MCP	13_a	15_a	7_b	13_a	8.35*
Effort					
TOMM	0	0	0	0	
MCP	0	0	6	0	
Learning					
TOMM	1	0	1	0	
MCP	1	0	2	1	
Atten/Conc					
TOMM	2	1	0	1	
MCP	1	0	0	1	
Mental					
Speed	0	0	0	0	
TOMM	0	0	0	0	
MCP					
Coordination	0	0	0	0	
TOMM	0	0	0	0	
MCP					
Motor Speed	0	0	0	0	
TOMM	0	0	0	0	
MCP					
Other	1	0	1	0	
TOMM	0	0	0	0	
MCP					
Don't Know	0	0	0	0	
TOMM	0	0	0	0	
MCP					

Note. Numbers with a * indicate that groups in that row are significantly different at $p <.05$. Rows with no Chi-Square Pearson coefficient had at least one variable that was a constant.

Performance on Effort Measures

Memory for Complex Pictures (MCP). The means and standard deviations for the number of correctly answered items on the MCP for Trials 1 and 2 are presented in Table 3. A two-way fixed effects ANOVA found significant group differences on MCP Trial 1, $F(3, 60) = 51.24$, $p<.01$, with Early and Late Warning Controls obtaining significantly higher scores on Trial 1 compared to both Early Warning CBIS and Late Warning CBIS, whose scores did not differ significantly from each other (*eta squared* = .47; large effect). No main effect for Warning group was found on Trial 1.

A two-way fixed effects ANOVA found significant group differences on MCP Trial 2, $F(3, 60) = 42.17$, $p<.01$, with Early and Late Warning Controls obtaining significantly higher scores on Trial 2 compared to both Early Warning CBIS and Late Warning CBIS, whose scores did not differ significantly from each other (*eta squared* = .42; large effect). No main effect for Warning group was found on Trial 2 (See Table 3).

Test of Memory Malingering (TOMM). The means and standard deviations for the number of correctly answered items on the TOMM for Trials 1 and 2 are presented in Table 3. A two-way fixed effects ANOVA revealed significant group differences on TOMM Trial 1, $F(3, 60) = 45.94$, $p<.01$, with Early and Late Warning Controls obtaining significantly higher scores on Trial 1 compared to both Early and Late Warning CBIS, whose scores did not differ significantly from each other (*eta squared* = .44; large effect). No main effect for Warning group was found on Trial 1.

A two-way fixed effects ANOVA revealed significant group differences on TOMM Trial 2, $F(3, 60) = 37.70$, $p<.01$, with Early and Late Warning Controls obtaining significantly higher scores on Trial 2 compared to both Early and Late Warning CBIS, whose scores did not differ significantly from each other (*eta squared* = .39; large effect). No main effect for Warning group was found on Trial 2 (See Table 3).

This data indicates that, as expected, the CBIS groups performed significantly lower than the Control groups. In addition, the timing of warning did not significantly affect performance on the tests of effort.

Table 3
Mean Correct Responses on the MCP

	Group				
	Early Warning Controls (n=15)	*Late Warning Controls (n=15)*	Early Warning *CBIS* (n=15)	Late Warning CBIS *(n=15)*	*F* (3, 60)
MCP -T1 *M (SD)*	48.93_a *(1.79)*	48.20_a *(3.10)*	31.73_b *(9.83)*	36.20_b *(11.84)*	51.24*
MCP -T2 *M (SD)*	49.20_a *(1.08)*	49.00_a *(2.07)*	33.73_b *(9.083)*	38.33_b *(12.45)*	42.17*

Note. Means in the same row having the same subscript are not significantly different at $p < .01$. T1 = Trial 1, T2 = Trial 2; *$p < .01$.

Table 4
Mean Correct Responses on the TOMM

	Group				
	Early Warning Controls (n=15)	*Late Warning Controls (n=15)*	Early Warning *CBIS* (n=15)	Late Warning CBIS *(n=15)*	*F* (3, 60)
TOMM-T1 *M (SD)*	48.27_a *(1.94)*	47.73_a *(3.86)*	33.53_b *(8.94)*	36.93_b *(10.69)*	45.94*
TOMM-T2 *M (SD)*	50.00_a *(0.00)*	49.40_a *(2.32)*	35.27_b *(10.22)*	39.67_b *(11.32)*	37.73*

Note. Means in the same row having the same subscript are not significantly different at $p < .01$. T1 = Trial 1, T2 = Trial 2; *$p < .01$.

Questionnaire Responses for Standard Tests

Face Validity Questionnaire (FVQ). Results from the FVQ were examined to determine the face validity of the standard measures. Participants' responses were categorized for each test based on previous research (DenBoer, 2007): (a) Memory, (b) Attention/Concentration, (c) Mental Speed, (d) Psychomotor Coordination, (e) Learning, (f) Effort, (g) Motor Speed, (h) Other, and (i) Don't Know.

Table 5 presents the frequencies of participant's beliefs regarding the nature of the standard tests. Frequencies for the two Early Warning groups represent participant's beliefs regarding the nature of the standard tests when receiving information that some of the tests may be tests of effort *prior* to administration of the tests. Frequencies for the Late Warning groups represent participant's beliefs regarding the nature of the standard tests when receiving information that some of the tests may be tests of effort *after* completing the tests.

A Pearson Chi-Square revealed that significant group differences for participants endorsement of Trails A as a test of mental speed, χ^2 (3, 60) = 9.01, p< .05 (*Cramer's V*= .37; medium effect). Posthoc Chi-Square revealed that individuals in the Early Warning CBIS and Controls were less likely to endorse Trails A as a test of mental speed than individuals in Late Warning Controls and CBIS. No other significant group differences were found. Regardless of group, participants primarily endorsed Finger Tapping as a test of Psychomotor Coordination and Motor Speed. Participants were unsure about the nature of Trails B and primarily endorsed it as Other, Attention/Concentration, and Mental Speed, in that order. Participants across groups endorsed Digit Span as a test of Memory. Digit Symbol-Coding was primarily endorsed as a test of Memory, but was also endorsed as a test of Mental Speed, Attention/Concentration, and Other. In summary, participants were fairly accurate in attributions of cognitive domain for the standard measures.

Table 5

Frequencies of Participants Endorsing Cognitive Domains on the FVQ –Standard Tests

Cognitive Domain	Group				
Measure	*Early Controls (n=15)*	*Late Controls (n=15)*	Early *CBIS* (n=15)	Late CBIS *(n=15)*	χ^2
Memory					
FT	0	0	1	1	
T-A	1	0	1	0	
T-B	0	0	2	2	
DS	8	10	9	9	3.70
DS-C	1	1	1	1	
Learning					
FT	1	0	0	1	
T-A	0	0	1	0	
T-B	0	0	0	0	
DS	0	2	0	0	
DS-C	4	2	3	2	
Atten/Conc					
FT	1	1	3	3	
T-A	3	1	1	2	
T-B	5	3	3	4	
DS	0	3	0	2	
DS-C	1	1	1	2	
Mental Speed					
FT	0	1	0	0	
T-A	7_a	12_b	5_a	10_b	12.20*
T-B	4	5	3	0	
DS	0	0	0	0	
DS-C	3	3	1	3	
Coordin					
FT	6	7	3	6	4.51
T-A	2	0	0	3	
T-B	1	1	0	0	
DS	0	0	0	0	
DS-C	0	0	1	2	
Motor Speed					
FT	4	5	8	3	5.88
T-A	2	2	4	0	
T-B	0	1	3	4	
DS	0	0	0	0	
DS-C	0	0	0	0	
Effort					
FT	0	0	0	0	
T-A	0	0	0	0	
T-B	0	0	0	0	
DS	0	0	2	0	
DS-C	0	0	0	0	

Other					
FT	2	1	0	1	
T-A	0	0	3	0	
T-B	3	5	4	5	
DS	6	0	2	4	
DS-C	6	8	8	4	1.69
Don't Know					
FT	1	0	0	0	
T-A	0	0	0	0	
T-B	2	0	0	0	
DS	2	0	0	0	
DS-C	0	0	0	0	

Note. Numbers with a * indicate that groups in that row are significantly different at p <.05. Rows with no Chi-Square Pearson coefficient had at least one variable that is a constant. DS – C = Digit Symbol – Coding; DS = Digit Span; FT = Finger Tapping; T – A = Trails A; T – B = Trails B.

Performance on Standard Measures

WAIS-III Digit Symbol-Coding. The means and standard deviations for the number of correctly answered items on Digit Symbol-Coding are shown in Table 6. A two-way fixed effects ANOVA revealed significant differences for Digit Symbol – Coding, $F(3, 60) = 65.64$, $p<.01$, with Early and Late Warning Controls Scoring significantly higher than Early and Late Warning CBIS (*eta squared* = .49; large effect). Additionally, Late Warning CBIS scored significantly higher (or more poorly) than Early Warning CBIS, $F(3, 60) = 8.38$, $p<.01$ (*eta squared* = .06; small effect).

WAIS-III Digit Span. The means and standard deviations for the number of correctly remembered items on Digit Span is shown in Table 6. A two-way fixed effects ANOVA revealed significant group differences on Digit Span – Total, $F(3, 60) = 5.11$, $p<.05$, with Early and Late Warning Controls scoring significantly higher than Early and Late Warning CBIS (*eta squared* = .18; medium effect). No main effect for Warning group was found.

Finger Tapping. The means and standard deviations for the average number of taps on Finger Tapping is shown in Table 6. A two-way fixed effects ANOVA revealed significant group differences on Finger Tapping, $F(3, 60) = 7.74$, $p<.01$, with Early and Late Warning Controls scoring significantly higher than Early and Late Warning CBIS (*eta squared* = .12; medium effect). No main effect for Warning was found.

Trails A. The means and standard deviations for the time to complete Trails A is shown in Table 6. A two-way fixed effects ANOVA revealed significant group differences on Trails-A, $F(3, 60) = 25.31$, $p<.01$, with Early and Late Warning Controls taking significantly less time than Early and Late Warning CBIS (*eta squared* = .29; large effect). An interaction revealed significant differences with Early Warning CBIS taking significantly longer than Late Warning CBIS, $F(3, 60) = 4.13$, $p<.05$ (*eta squared* = .05; small effect).

Trails B. The means and standard deviations for the time to complete Trails B on is shown in Table 6. A two-way fixed effects ANOVA revealed significant group differences on Trails-B, $F(3, 60) = 14.76$, $p<.01$, with Early and Late Warning Controls taking significantly less time than Early and Late Warning CBIS (*eta squared* = .18; medium effect). An interaction revealed significant differences with Early Warning CBIS taking significantly longer than Late Warning CBIS, $F(3, 60) = 7.73$, $p<.01$ (*eta squared* = 09.; medium effect).

In summary, as expected, Controls performed significantly better than CBIS on all standard tests. On Digit Symbol-Coding and Trails A and B, Late Warning CBIS performed significantly better than Early Warning CBIS. Warning did not have a significant effect for Digit Span and Finger Tapping.

Table 6
Mean Performance on Standard Neuropsychological Measures

	Group				
Measure	*Early Warning Controls (n=15)*	*Late Warning Controls (n=15)*	Early Warning *CBIS* (n=15)	Late Warning CBIS *(n=15)*	*F* (3, 60)
DS-C M*(SD)*	86.47_a *(11.88)*	81.80_a *(16.85)*	43.79_c *(15.88)*	60.13_b *(16.33)*	65.64*
DS M*(SD)*	16.00_a *(4.19)*	15.60_a *(2.75)*	10.86_b *(4.66)*	13.44_b *(3.54)*	12.80*
FT M*(SD)*	301.67_a *(93.55)*	324.80_a *(164.96)*	210.07_b *(173.69)*	211.44_b *(127.45)*	8.74*
T-A M*(SD)*	23.23_a *(8.64)*	25.77_a *(13.98)*	57.90_b *(25.08)*	40.50_c *(22.00)*	25.31*
T-B M*(SD)*	52.09_a *(13.29)*	57.20_a *(23.21)*	148.69_b *(102.63)*	72.24_c *(25.63)*	14.76*

Note. Means in the same row having the same subscript are not significantly different at $p < .05$. DS – C = Digit Symbol – Coding; DS = Digit Span; FT = Finger Tapping; T – A = Trails A; T – B = Trails B. *$p < .01$.

Manipulation Check

Manipulation Check Questionnaire (MCQ). The means and standard deviations for how hard participants tried and success ratings are presented in Table 7. All participants whose data was included in the study endorsed adequate adherence to their participant instructions. A manipulation check questionnaire was administered to all participants after they had completed all neuropsychological tests. The MCQ contained three questions designed to make certain that participants followed their instructions accurately and tried adequately hard to follow their experimental instructions. "Adequate" was defined by the participants indicating a 5 or higher on the Likert-scale for all three questions. All participants included in this study's analysis were able to recall their instructions accurately and indicated that they had, at the very least, tried moderately hard to follow their experimental instructions during the study. All participants included in the analysis also indicated they were at least moderately successful in carrying out their experimental instructions.

MCQ ratings were analyzed by three separate one-way ANOVAs. Significant group differences were found for participants indication of trying to the best of their ability on the measures, $F(3, 60) = 17.13$, $p<.01$. Tukey HSD comparisons revealed that Early and Late Warning Control participants rated their attempt to perform well on the tests as significantly greater than Early and Late Warning CBIS, which did not differ significantly from each other (Table 5).

Significant group differences were found for how much hard participants tried to follow their instructions, $F(3, 60) = 3.11$, $p<.05$. Tukey HSD comparisons revealed that participants in Late Warning CBIS rated how hard they tried to following their instructions as significantly greater than Late Warning Controls. All other groups did not differ significantly from one another.

Significant group differences were found for how successful participants felt they were in following their instructions, $F(3, 60) = 20.60$, $p<.01$. Tukey HSD comparisons revealed that Early and Late Controls rated their success at following their instructions significantly greater than Early and Late CBIS, who did not differ significantly from one another.

Table 7
Mean Attempt on Measures, Attempt at Following Directions, and Success at Following Directions

	Group				
Ratings	*Early Warning Controls (n=15)*	*Late Warning Controls (n=15)*	Early Warning *CBIS* (n=15)	Late Warning CBIS *(n=15)*	*F* (3, 60)
Tried to Best of Ability M*(SD)*	9.00_a *(1.41)*	9.60_a *(0.51)*	7.71_b *(1.68)*	6.69_b *(1.08)*	17.13*
Tried to Follow Instructions M*(SD)*	9.27 *(1.44)*	9.60_a *(0.63)*	8.79 *(0.98)*	8.44_b *(1.32)*	3.11*
Success M*(SD)*	8.80_a *(1.70)*	8.40_a *(1.18)*	6.21_b *(0.70)*	6.00_b *(1.16)*	20.60*

Note. Means in the same row having the same subscript are not significantly different at $p < .05$ in the Tukey HSD comparison. 1 = low effort/success in following instructions, 10 = high effort/success in following instructions.

Experiment 2

Minimal research on the face validity of effort tests exists. Many effort tests include a feedback mechanism that provides information on how well the examinee is performing during the test. How feedback influences the performance of examinees that are providing inadequate effort is unclear. Feedback may alert the examinee to the true nature of the test and aid the individual in avoiding detection (Bolan, Foster, Schmand, & Bolan, 2002). On the other hand, feedback may actually cause the examinee to perform worse. For example, participants will hear that their answers are correct at least half of the time unless they are providing inadequate effort (Lezak, 1995). In individuals who are attempting to appear impaired, this may encourage participants to worsen their performance (Tombaugh, 1997). Because it is unclear how feedback influences test performance, it is important that the effects of feedback are clarified and empirically supported. Due to the importance of face validity for effort tests as discussed previously, a study examining the effect of feedback on face validity was conducted.

Purpose of Experiment 2

The purpose of the second experiment was to determine the effect of feedback (visual or no feedback) on the performance of a newly developed test of effort, the Memory for Complex Pictures (MCP; DenBoer & Hall, 2005; DenBoer, 2007). The MCP was chosen because it is a measure that provides an option of different feedback modes (visual or no feedback).

Hypotheses for Experiment 2

1. It was hypothesized that participants in the Control group would score significantly higher on the MCP compared to participants in the CBIS group with no feedback.
2. It was hypothesized that the participants in the CBIS group with no feedback would score significantly higher on the MCP compared to participants in the CBIS group with feedback.

Method for Experiment 2

Participants

Sixty-five University of Montana undergraduate psychology students were recruited to participate in this study. Students were 18 years of age or older and were excluded from the study if they reported a history of psychological or neurological problems or treatment. Participants were divided into four groups, Feedback Control, No Feedback Control, Feedback Coached Brain Injury Simulator (CBIS), or No Feedback

CBIS. The 65 students were given 3 credits toward their respective courses for participation. Each student completed the Medical and Health History Questionnaire, the MCP, and the Face Validity Questionnaire.

Materials

Medical and Health History Questionnaire. A medical and health history questionnaire was used as a screening measure to assess participants' history of psychological, neurological, and substance abuse problems (See Appendix A). As discussed in Experiment 1.

Memory for Complex Pictures (MCP). The MCP is a forced-choice, two-alternative measures consisting of complex digital photographs for assessing effort during neuropsychological evaluation. A full discussion of the measure is in Experiment 1. For Experiment 2, the different feedback options were changed for the Feedback and No Feedback groups. For the Feedback group, during the recognition trial, if the examinee chose the target stimulus the word "RIGHT" appeared in all caps, 18-point, bold font, approximately two inches to the right of the target stimulus. Both the target stimulus and the foil were present for .75 seconds. If the examinee selected the foil instead of the target, the word "WRONG" appeared to the right of the foil and both pictures remained on screen for .75 seconds. In the No Feedback group, no visual or auditory feedback was provided to the participant.

Manipulation Check Questionnaire (MCQ). Designed as a "manipulation check," the MCQ contained three questions designed to make certain all participants included in the study's analysis were able to recall their instructions accurately and indicated that they had put forth at least moderate effort following their experimental instructions during the study (See Appendix C). A full discussion of the questionnaire is in Experiment 1.

Procedure

Participants were brought into the lab and provided one of two types of instructions. The order of these instructions was randomized. Control participants were asked to perform to the best of their ability. Coached Brain Injury Simulators (CBIS) received the following instructions:

> You are about to take a cognitive test that may examine mental abilities such as attention, memory, thinking, and reasoning skills, and your ability to think quickly.
> While responding to the test, please pretend that you have experienced brain damage from a car accident involving a head-on collision. You hit your head against the windshield and were knocked out for 15 minutes. Afterwards, you felt "dazed" so you were hospitalized overnight for observation. Because the driver of the other car is at fault, you have decided to go to court to get money from the person responsible. During the next few months following the accident, the negative effects from your head injury disappear. Your lawsuit has not been settled yet, and your lawyer has told you that you may get more money if you look like you are suffering from brain damage.
> As you pretend to be this car accident victim, try to respond to each test as a patient who is trying to appear brain damaged in order to get money from the lawsuit. Thus, your performance on the tests should convince the examiner as well as the people involved in deciding the outcome of your lawsuit that you are still suffering from brain damage. In order to convince these individuals, your brain damage must be believable. *If you succeed in convincing the examiner that you have suffered a brain injury, you will receive two additional experimental credits, for a total of 3 credits.*
> Try to produce the most severe problems that you can *without* making it too obvious to the examiner. Major exaggerations, such as remembering absolutely nothing, are easy to detect. If the examiner does not believe that you have any problems you will not get any money for your head injury. People who have a head injury often have

> problems paying attention, cannot remember things as well, and do not learn things as easily as they did before their injury. They also think a little slower than they used to. Keep this in mind when taking the tests. Remember you are to try to mimic the performance of persons who are truly brain damaged.

Participants in the Control and CBIS groups were also randomly assigned to one of two feedback groups (Feedback or No Feedback). In the Feedback group, participants saw the word "RIGHT" next to the picture they accurately identified on the MCP or "WRONG" next to the picture they inaccurately identified. In the No Feedback group, participants were not presented with feedback.

All participants were administered the medical and health history questionnaire as a screening measure prior to participation in the study. After receiving the scenario, the MCP was administered.

Results for Experiment 2

Power for Experiment 2

For the purposes of this analysis power was set at .80 with a significance level of $\alpha = .05$ (Cohen, 1992). Large effect sizes were assumed based on the work of Huskey (2005) and DenBoer (2007), whose results with similar measures yielded R-squared effect size estimates for type of instructions (i.e., group) exceeding .40 and power ranging from .92 to 1.0. Given that mean differences were collected, 15 participants per group were required for the analyses. Provided that a 2 x 2 ANOVA was required to be conducted, a minimum of 60 total participants were needed in Experiment 1.

Demographic Information

A total sample size 65 undergraduate students completed the battery of neuropsychological measures. Demographic characteristics are shown in Table 8. No significant differences for gender were found between the four groups, χ^2 (3, 65) = .149, *ns*. Group differences for Age and Education were analyzed using two separate one-way ANOVAs. There were no significant differences found for Age, *F*(3, 65) = 2.03, *ns* or Education, *F* (3, 65) = 2.00, *ns*.

Table 8
Demographic Information for Participants in Experiment 2

	Group				
	Feedback Controls *(n=15)*	No Feedback Controls *(n=17)*	Feedback CBIS *(n=14)*	No Feedback CBIS *(n=17)*	χ^2 *or F*
Gender					
Males (n)	6	7	5	5	.149
Females (n)	9	10	9	12	
Age					
M	24.13	20.24	20.93	20.65	2.03
(SD)	*(7.05)*	*(2.56)*	*(5.11)*	*(4.27)*	
Education					
M	13.80	12.82	12.57	13.06	2.00
(SD)	*(2.11)*	*(0.88)*	*(1.16)*	*(1.39)*	

Note. Means in the same row having different subscripts are significantly different at $p<.05$ in the Tukey HSD comparison. * $p<.05$

Performance on Neuropsychological Measures

Memory for Complex Pictures (MCP). Results for performance on the MCP Trial 1 and 2 is shown in Table 9. A two-way fixed effects ANOVA on MCP Trial 1 revealed no significant main effect for Feedback. A significant main effect for Group was found, *F*(3, 65) = 52.95, $p<.05$, with Feedback and No Feedback Controls scoring significantly higher than Feedback and No Feedback CBIS (R squared = .49; *large effect*). No significant interaction was found.

A two-way fixed effects ANOVA on MCP Trial 2 revealed no significant main effect for Feedback. A significant main effect for Group was found, $F(3, 65) = 39.76$, $p<.05$, with Feedback and No Feedback Controls scoring significantly higher than Feedback and No Feedback CBIS (R squared = .43; *large effect*). No significant interaction was found.

Table 9
Mean Correct Responses on the MCP for Experiment 2

	Group				
	Feedback Controls (n=15)	*No Feedback Controls (n=15)*	Feedback *CBIS* (n=15)	No Feedback CBIS *(n=15)*	*F* (3, 65)
MCP -T1	49.47_a	48.35_a	34.64_b	29.76_b	52.95*
M (SD)	*(1.36)*	*(4.30)*	*(9.79)*	*(14.36)*	
MCP -T2	49.80_a	48.53_a	37.29_b	31.06_b	39.76*
M (SD)	*(0.41)*	*(4.54)*	*(10.81)*	*(14.50)*	

Note. Means in the same row having the same subscript are not significantly different at $p < .01$. T1 = Trial 1, T2 = Trial 2; $*p < .001$.

Questionnaire Results for Experiment 2

Manipulation Check Questionnaire (MCQ). The means and standard deviations for how hard participants tried on the measures and success ratings are presented in Table 10. All participants whose data was included in the study endorsed adequate adherence to their participant instructions. A manipulation check questionnaire was administered to all participants after they had completed all neuropsychological tests. The MCQ contained three questions designed to make certain that participants followed their instructions accurately and tried adequately hard to follow their experimental instructions. “Adequate” was defined by the participants indicating a 5 or higher on the Likert-scale for all three questions. All participants included in this study’s analysis were able to recall their instructions accurately and indicated that they had, at the very least, tried moderately hard to following their experimental instructions during the study. All participants included in the analysis also indicated they were at least moderately successful in carrying out their experimental instructions.

MCQ ratings were analyzed by three separate one-way ANOVAs. Significant group differences were found for participants indication of trying to the best of their ability on the measures, $F(3, 65) = 10.22$, $p<.01$. Tukey HSD comparisons revealed that Feedback and No Feedback Control participants rated how hard they tried to do well on the measure as significantly greater than Feedback and No Feedback CBIS, who did not differ significantly from each other (Table 10).

Significant group differences were found for how hard participants tried to follow their instructions, $F(3, 65) = 10.51$, $p<.01$. Tukey HSD comparisons revealed that participants in the Feedback and No Feedback Control group rated how hard they tried to follow instructions significantly greater than Feedback and No Feedback CBIS group, who did not differ significantly from each other (Table 10).

Significant group differences were found for how successful participants felt they were in following their instructions, $F(3, 65) = 30.62$, $p<.01$. Tukey HSD comparisons revealed that Feedback and No Feedback Controls rated their success at following their instructions significantly greater than Feedback and No Feedback CBIS, who did not differ significantly from one another (Table 10).

Table 10

Mean Attempt on Measures, Attempt at Following Directions, and Success at Following Directions on Experiment 2

	Group				
Ratings	*Feedback Controls (n=15)*	*No Feedback Controls (n=15)*	Feedback *CBIS* (n=15)	No Feedback CBIS *(n=15)*	*F* (3, 60)
Tried to Best of Ability M*(SD)*	9.53_a *(0.92)*	9.18_a *(1.07)*	6.64_b *(2.73)*	7.59_b *(1.87)*	10.22*
Tried to Follow Instructions M*(SD)*	9.33_a *(0.98)*	9.65_a *(0.61)*	7.00_b *(2.35)*	7.94_b *(1.56)*	10.51*
Success M*(SD)*	9.73_a *(0.59)*	9.59_a *(0.87)*	6.21_b *(0.70)*	6.90_b *(1.81)*	30.62*

Note. Means in the same column having the same subscript are not significantly different at $p < .05$ in the Tukey HSD comparison. 1 = low effort/success in following instructions, 10 = high effort/success in following instructions.

Discussion

Experiment 1 was designed to examine the role of a warning about the presence of effort tests on face validity and performance on a neuropsychological battery. Sixty undergraduate students participated and only those who did not report a history of neurological or psychological problems were included in the analysis. A significant difference for age was found between Early Warning Controls and Late Warning Controls. However, review of the raw data indicated that this difference is the result of one significantly older participant and it is highly unlikely this would have an effect on the results.

Face Validity of Effort Tests

The MCP and TOMM were compared in terms of face validity, as high face validity has been shown to be an important trait in a successful SVT (e.g., DenBoer, Hall, Jacobsen, and Hoffman, 2006; Huskey, 2005; Huskey & Hall, 2003; Tan, Slick, Strauss, & Hultsch, 2003). The original hypothesis that the MCP would demonstrate improved face validity as a genuine measure of memory when compared to the TOMM was not supported. The MCP and TOMM were found to demonstrate relatively equivalent face validity across all participant groups. It is notable, however, that the newly developed MCP demonstrated essentially equal overall face validity to the most commonly used effort test (Sharland & Gfeller, 2007; Slick, Tan, Strauss, & Hultsch, 2004). These results add to the undersized literature base examining the face validity of neuropsychological measures, and, in particular, of effort measures (Huskey, 2005; Huskey & Hall, 2003; Kafer & Hunter, 1997; Rees, Tombaugh, Gansler, & Moczynski, 1998; Tan, Slick, Strauss, & Hultsch, 2003; Tombaugh, 1997). This finding lends support to the continued research efforts with the MCP as a computerized measure of effort.

Across groups, participants were equally likely to view the TOMM as a test of memory. It does not appear that group or timing of warning influenced the perception of the TOMM. However, the MCP was viewed as a test of memory less frequently in the Early Warning CBIS group compared to all three other groups. Additionally, and of great interest, is that, as hypothesized, more Early Warning CBIS participants endorsed the MCP as a test of effort than both the Late Warning CBIS and Control groups. In other words, it appears that providing a warning prior to administration of the effort tests made the participants more suspicious of the effort tests than if the warning was presented after the administration of the tests or not at all. This result is consistent with previous research indicating that warning participants prior to testing about the presence of measures designed to detect inadequate effort can challenge the accuracy of the assessments (Johnson & Lesniak-Karpiak, 1997; Slick, Tan, Strauss, & Hultsch, 2004; Youngjohn, Lees-Haley, and Binder, 1999). It is important to consider the implications of reduced face validity due to a warning. Slick, Tan, Strauss, and Hultsch (2004) found that almost 38% always provide a warning prior to testing. The results of the current study suggest that providing a warning prior to testing may lessen the face validity of the effort measures and may challenge the accuracy of the assessment.

Regardless of when the warning was delivered, both the MCP and TOMM were primarily endorsed as an actual measure of visual memory. Contrary to the original hypothesis, the MCP and TOMM were not endorsed as tests of effort more frequently in the Late Warning CBIS group than the Control groups. In fact, no participants in the Late Warning CBIS group endorsed the MCP or TOMM as a test of effort. This suggests that participants are not likely to be suspicious of a test when asked to evaluate the face validity of the measure after having taken it.

Interestingly, the TOMM was only endorsed as a test of effort in one of the groups (7%) and not significantly more frequently than the other groups. The MCP was viewed as a test of effort by 29% of the Early Warning CBIS participants, significantly more than the other three groups. This difference may be due to the fact that the MCP's visual stimuli include pictures of familiar locations and landmarks surrounding the Missoula area. Notably, research assistants reported that several participants specifically commented on their familiarity with some of the pictures in the measure. The familiarity of the participants with the MCP test stimuli may have increased suspicion regarding the nature of the test and produced inflated results for the MCP. This suggests that while the MCP was endorsed as a test of effort by more participants in one group, this result may be inflated. Caution should be used when generalizing these results outside of the Missoula area and it is highly suggested that this study be replicated outside of the area.

Performance on Effort Tests

As hypothesized, Control participants demonstrated near-ceiling performance on the MCP and TOMM. These scores were significantly higher than those of the CBIS groups. The finding that controls performed at near-ceiling levels on the TOMM and MCP is in line with results obtained by DenBoer (2007), Huskey (2005), and Tombaugh (1996; 1997). Contrary to the original hypothesis, Early Warning CBIS did not obtain significantly higher scores on the MCP and TOMM than the Late Warning CBIS on either the TOMM or the MCP even with reduced face validity. This has important implications for neuropsychological assessments. Providing a warning prior the administration of effort testing may in fact lessen the face validity of the effort test. The original hypothesis was that increased suspicious would lead to less exaggeration and therefore increased performance. In other words, it was thought that increased suspicion would create more sophisticated test takers who performed at higher levels than simulators who were not given a warning. However, although the results indicate that participants in the Early Warning CBIS were more suspicious of the effort tests and more likely to see them as tests of effort, they did not improve their performance as hypothesized. Indeed, the timing of the warning did not have a positive effect on the performance of the Early Warning CBIS participants.

It is unclear why simulators who were given an early warning would be more suspicious of a measure and also perform more poorly. However, for simulators who have little familiarity with the effects of brain injury, suspicion that a test may in fact detect their faking may encourage them to attempt to appear even more injured. In this way, an early warning may have decreased performance on the measures in an attempt to appear more authentic and less detectable. This finding is consistent with research by Heaton, Smith, Lehman, and Vogt (1978) that found that volunteer malingerers (simulators) are likely to demonstrate exaggerated poor performance on neuropsychological tests. This is an important finding because it suggests that providing an early warning does not necessarily dissuade potential malingerers from faking. In fact, an early warning may actually produce more exaggerated poor performance, making potential malingerers *more* detectable. In order to know how an early warning would influence neuropsychological performance in a clinical population, it would be important to replicate this study using a neurological and/or forensic population in order to see whether or not the same effect is found.

This study did not examine the influence of warning on a group of uncoached participants. Whereas the simulators in this study were told that head injured individuals often perform more poorly on memory tests and advised not to overexaggerate, uncoached participants are given less information on common cognitive difficulties associated with a head injury. It may be that uncoached simulators may have underperformed to an even greater degree given a warning that some tests may detect head injury. However, the salience of both the coaching instructions as well as the warning is unclear. Because the coaching and warning are imbedded in the scenario (see

Appendix F), we cannot be certain how closely participants are reading or remembering what they are instructed to read. Further research could look at how a warning influences the performance of an uncoached group as well as examining more closely how salient both the coaching instructions and the warning are to the participants.

Face Validity of Standard Tests

Contrary to the original hypothesis, timing of warning did not appear to have an effect on the face validity of the standard tests. Regardless of group, participants were primarily viewing the standard tests as belonging to one of the appropriate primary cognitive domains. Finger Tapping was primarily viewed a test of Psychomotor Coordination and Motor Speed. Digit Span and Digit-Symbol Coding were primarily endorsed as a test of Memory. Surprisingly, Trails A was primarily viewed as a test of Mental Speed, although Early Warning Controls and CBIS endorsed it as a test of Mental Speed more frequently than Late Warning Controls and CBIS, who endorsed Trails A in more varied cognitive domains. It appears that timing of warning may have had an effect on the face validity of Trails A, although the exact nature of the influence is unclear. It may be because Trails A is a more simple task which may appear “too easy” to participants. Trails B was primarily endorsed as Other, Attention/Concentration, and Mental Speed. This may be related to the complexity of the cognitive demands on Trails B.

A limitation of the study is the use of previously determined categories. Further examination of the categories suggested the use of fewer categories. For instance, Learning and Memory are separate categories as suggested by DenBoer (2007). However, both Learning and Memory essentially represent very similar cognitive domains. Similarly, a collapse of the categories for Psychomotor Coordination and Motor Speed is indicated in future studies. Percentages for individual cognitive domains may be somewhat lower due to the separation of essentially identical cognitive categories.

Performance on Standard Tests

As hypothesized, Control participants performed significantly better on the standard measures than CBIS participants. These results are in line with previous research by DenBoer (2007) and Huskey (2005). The original hypothesis suggested that Early Warning groups should perform better because they are more suspicious of the measures' ability to detect faking and thus less likely to overexaggerate their impairments. In other words, receiving a warning prior to administration of the measures was hypothesized to make participants more suspicious of the intent of the measure. This lowered face validity was hypothesized to make the participants more sophisticated in their test taking approach. However, participants in the Late Warning CBIS group performed significantly better on Digit Symbol-Coding and Trails A and B compared to Early Warning CBIS. This result was not found for Digit Span and Finger Tapping. Interestingly, receiving the warning prior to administration of the tests decreased the participant's performance on the measures. This finding may indicate that the early warning actually worked to increase the need to appear brain injured and made participants more likely to exaggerate. In other words, increased suspicion created by the early warning may have suggested to CBIS participants that they need to appear increasingly injured to avoid detection. In doing so, participants may have underperformed to such a great extent that they naively increased their ability to be detected. The late warning CBIS group, however, was not provided with increased suspicion and their performance may reflect this.

Performance on Effort Tests with and without Feedback

Given that little research has examined the role of feedback on performance on effort tests, Experiment 2 was designed to explore the influence of feedback on controls versus simulating participants. As expected, participants in the Control group performed significantly better than participants in the CBIS group. However, no effect was found for the presence of feedback. Contrary to the original hypothesis, participants in the Control group did not score significantly better than participants in the No Feedback CBIS group. Additionally, participants in the No Feedback CBIS group did not perform significantly better than participants in the Feedback CBIS group. Overall, it appears that the presence of visual feedback does not have an effect on performance even for participants who are faking impairment. These findings are in line with previous research by Bolan, Foster, Schmand, and Bolan (2002) who conducted three experiments to validate the use of the English language version of the Amsterdam Short Term Memory Test (ASTM test), a malingering-detection measure. The researchers found that immediate feedback on the accuracy of test responses had no significant effect on performance. However, the researchers did note trends in the direction of statistical significance, with the presence of immediate feedback influencing patients to perform worse on both the ASTM test and the TOMM.

The results of Experiment 1 emphasize the challenging nature of effort detection and push for further exploration into factors that may affect face validity and test performance. This experiment provided some evidence for an influence of warning simulators about the presence of effort tests on the face validity of effort measures. Specifically, the participants who were asked to simulate head injury were more likely to see both standard tests and effort tests and measures of effort if they were given a warning prior to administration of the measures. However, the nature of this influence is not entirely clear, and although possible reasons have been suggested here, further research must be done to fully understand how the warning is impacting face validity as well as test performance. The results from Experiment 1 indicate that although participants in the early warning simulator group were more readily seeing the measures as tests of effort, they were actually performing more poorly than participants who were given the warning after the administration of the measures. It is not fully clear why participants would be more suspicious of a test but not alter their performance to avoid detection. Further research examining how face validity affects test performance is indicated.

The results from Experiment 2 call into question the use of feedback during effort testing. Little empirical research exists regarding this topic and the few studies that have been conducted have produced similar results to the ones in this study. Overall, it appears that feedback may not produce more adequate effort or easier detection. Further research should be done to elucidate the influence, if any, of feedback on effort testing performance. It would be particularly important to examine the effects of feedback using a neurological population instead of using healthy controls as simulators.

Conclusion

The results of the current study demonstrated that the timing of warning does appear to play a role in the face validity of effort tests. Notably, a group of simulators who were provided with a warning about the presence of one effort test prior to administration of the tests were less likely to view the effort test as tests of memory and more likely to view them as actual tests of effort. Although the warning had an effect on face validity, it did not appear to have a clear effect on test performance. This suggests that in a simulating population, providing an early warning may not produce more sophisticated test takers. However, it is still unclear how the early warning would operate in a sample of neurological or forensic individuals. Future research is needed to better understand the nature of the relationship between warning and effort test face validity and performance. This research suggests using caution when providing a warning about the presence of effort tests during neuropsychological assessment. Additionally, the current study demonstrated that feedback during effort testing does not appear to influence performance on effort tests. If the use of feedback to deter inadequate effort continues to be unsupported by empirical data, this may suggest that feedback is an unnecessary component of effort testing. Further research is needed to understand the role of feedback and to potentially challenge the necessity of feedback in effort testing.

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed.,Text Revision (DSM-IV-TR). Washington, DC: American Psychiatric Association.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Ardolf, B.R., Denney, R.L., & Houston, C.M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *Clinical Neuropsychologist, 21*, 899-916.

Aronoff, G.M., et al. (2007). Evaluating malingering in contested injury or illness. *World Institute of Pain, 7,* 178-204.

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

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

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.

Bornstein, R.F., Rossner, S.C., Hill, E.L., & Stepanian, M.L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment, 63,* 363-386.

DenBoer, J.W. (2007). Memory for Complex Pictures (MCP): Development and validation of a digital test of effort (Doctoral dissertation, The University of Montana, 2007). *Dissertation Abstracts International, 68,* 3B.

DenBoer, J.W. & Hall, S. (2007). Neuropsychological test performance of successful brain injury simulators. *The Clinical Neuropsychologist, 21,* 943-955.

DenBoer, J.W. & Hall, S. (2007). Preliminary Validation of the Memory for Complex Pictures test in a Mixed Brain-Injured Population. Poster presented at the International Neuropsychological Society, Portland, OR, February.

Delain, S.L., Stafford, K.P., & Ben-Porath, Y.S. (2003). Use of the TOMM in a criminal court forensic assessment setting. *Assessment, 10,* 370-381.

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

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

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

Flaro, L., Green, P., & Robertson, E. (2007). Word memory test failure 23 times higher in mild brain injury than in patients seeking custody: The power of external incentives. *Brain Injury, 21,* 373-383.

Gervais, R.O., Green, P., Allen, L.M., & Iverson, G.L. (2001). Effects of coaching on Symptom Validity Testing in chronic pain patients presenting for disability assessments, *Journal of Forensic Neuropsychology*, *2,* 1–19.

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.

Gorny, I. & Merten, T. (2005). Symptom information—warning—coaching: How do they affect successful feigning in neuropsychological assessment? *Journal of Forensic Neuropsychology, 4,* 71-97.

Green, P. & Iverson, G.L. (2001). Validation of the computerized assessment of response bias in litigating patients with head injuries. *The Clinical Neuropsychologist, 15*, 492-497.

Green, P., Iverson, G.L. & Allen, L.M. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13,* 813-819.

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

Halstead, W.C. *Brain and Intelligence.* Chicago: University of Chicago, 1947.

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

Heilbrun, K., Bennett, W.S., White, A.J., & Kelly, J. (1990). An MMPI-based empirical model of malingering and deception. *Behavioral Sciences and the Law, 8,* 45-53.

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

Iverson, G.L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology, 13,* 77-90.

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

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,* 1094-1102.

Moore, B.A. & Donders, J. (2004). Predictors of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury, 18,* 975-984.

Nelson, T.O., Metzler, J. & Reed, D.A. (1974). Role of details in the long-term recognition of picture and verbal descriptions. *Journal of Experimental Psychology, 102*, 184-186.

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

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

Resnick, P.J. *Clinical Assessment of Malingering and Deception*. New York: Guilford

Press, 1997.

Rey, A. (1964). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie, 28,* 286-340.

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

Rose, F.E., Hall, S., Szalda-Petree, A.D., & Bach, P.J. (1998). A comparison of four tests of malingering and the effects of coaching. *Archives of Clinical Neuropsychology, 13,* 349-363.

Sharland, M.J. & Gfeller, J.D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22,* 213-223.

Slick, D.J., Tan, J.E., Strauss, E.H., & Hultsch, D.F. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology, 19,* 465-473.

Strauss, E., Sherman, E.M.S, & Spreen, O. (2006) A Compendium of neuropsychological tests: Administration, norms, and commentary. (3rd ed.). NY: Oxford University Press.

Suhr, J.A. & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology, 15,* 415-424.

Tan, J.E., Slick, D.J., Strauss, E., & Hultsch, D.F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist, 16,* 495-505.

Tombaugh, T.N. (1996). *TOMM. Test of Memory Malingering*. North Tonawanda: NY: Multi-health systems.

Tombaugh, T.N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9,* 260-268.

Weschler, D. (1997). *WAIS-III Administration and scoring manual*. San Antonio: Harcourt Brace & Company.

Wetter, M.W. & Corrigan, S.K. (1995). Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. *Professional Psychology: Research and Practice, 26,* 474-477.

Youngjohn, J.R., Burrows, L., & Erdal, K. (1995). Brain damage or compensation neurosis? The controversial post-concussive syndrome. *Clinical Neuropsychologist, 9,* 112-123.

Youngjohn, J.R., Lees-Haley, P.R., & Binder, L.M. (1999). Comment: Warning malingerers produces more sophisticated malingering. *Archives of Clinical Neuropsychology, 14,* 511-515.

Appendix A

Medical History Form

PLEASE FILL OUT THIS MEDICAL AND HEALTH HISTORY QUESTIONNAIRE

Date_______ Age_________ Sex _______ Race________ #_______

Were there any known difficulties with your birth? Yes No
If yes, describe____
Do you have a vision problem that requires corrective lenswear (e.g., glasses)? Yes No

Education
Did you ever have to repeat any grades? Yes No
Were you ever placed in special education classes? Yes No
What is the highest grade have you completed? _____
(e.g., if you are a college freshman you have completed 12 yrs. of ed.)

Medical and Health History

	Yes	No
1. Have you ever been diagnosed with any neurological condition? If so, please list: ____	___	___
2. Have you ever had a blow to your head in which you were unconscious for longer than 30 minutes?	___	___
3. Are you currently experiencing significant problems with your mood (anxiety and/or depression) or any other psychiatric condition? If so, please list: ____	___	___
4. Are you currently receiving treatment for your mood (anxiety and/ or depression) or any other psychiatric condition?	___	___
5. Have you ever felt you should cut down on your drinking/drug use?	___	___
6. Have you ever been annoyed by people that criticize your drinking/ drug use?	___	___
7. Have you felt bad or guilty about your drinking or drug use?	___	___
8. Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover?	___	___
9. Do you often drive under the influence of alcohol/drugs?	___	___

Appendix B

Role-Play Termination Instructions

If you have received instructions to pretend like you sustained brain damage, at this point in the study please **stop** following your instructions. From this point forward in the study please provide your personal and honest responses to all questions. Thank you.

Appendix C

MC Questionnaire

1. Please summarize the instructions you were given by the examiner at the beginning of this experiment:

____

____

2. Please rate the effort you put in to this study *given your instructions* to do the best you could on the measures in this study:

1	2	3	4	5	6	7	8	9	10
No effort at all				Moderate effort					Maximum effort

3. Indicate how hard you tried to follow the instructions you were given at the beginning of the experiment by circling the number that best describes your effort.

1	2	3	4	5	6	7	8	9	10
Didn't try at all				Tried moderately hard					Tried very hard

4. Indicate how successful you think were in producing the results asked of you in the instructions by circling the number that best describes your effort.

1	2	3	4	5	6	7	8	9	10
Not at all successful				Moderately successful					Very successful

Appendix D

FV Questionnaire

Please indicate what you think each test was designed to measure as specifically as possible.

5. What do you think the test with different numbers in circles (connected in dot-to-dot fashion) was designed to measure?

How certain are you about your answer?

1	2	3	4	5	6	7	8	9	10
Not at all certain				Somewhat certain					Very certain

6. What do you think the test with different *numbers* and *letters* in circles (connected in a dot-to-dot fashion) was designed to measure?

How certain are you about your answer?

1	2	3	4	5	6	7	8	9	10
Not at all certain				Somewhat certain					Very certain

7. What do you think the test with different numbers and symbols (the test the provided a key matching symbols with numbers) was designed to measure?

How certain are you about your answer?

1	2	3	4	5	6	7	8	9	10
Not at all certain				Somewhat certain					Very certain

8. What do you think the test that asked you to remember numbers forwards and backwards to trying to measure?

____

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10

Not at all certain — Somewhat certain — Very certain

9. What do you think the computer test that asked you to remember digital photographs was designed to measure?

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10

Not at all certain — Somewhat certain — Very certain

10. What do you think the test that asked you to remember line-drawings was designed to measure?

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10

Not at all certain — Somewhat certain — Very certain

11. What do you think the test that required you to tap your finger as fast as you could was designed to measure?

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10

Not at all certain — Somewhat certain — Very Certain

Appendix E

FV Questionnaire

The tests administered today test cognitive abilities such as memory, attention, and speed of information processing. It is possible that some of the tests you took today were designed to detect if someone is faking brain damage. Please indicate what you think each test was designed to measure as specifically as possible.

1. What do you think the test with different numbers in circles (connected in dot-to-dot fashion) was designed to measure?

____

____

____

____

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10
Not at all certain Somewhat certain Very certain

2. What do you think the test with different *numbers* and *letters* in circles (connected in a dot-to-dot fashion) was designed to measure?

____

____

____

____

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10
Not at all certain Somewhat certain Very certain

3. What do you think the test with different numbers and symbols (the test the provided a key matching symbols with numbers) was designed to measure?

____

____

____

____

How certain are you about your answer?

____

1 2 3 4 5 6 7 8 9 10
Not at all certain Somewhat certain Very certain

4. What do you think the test that asked you to remember numbers forwards and backwards to trying to measure?

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10
Not at all certain Somewhat certain Very certain

5. What do you think the computer test that asked you to remember digital photographs was designed to measure?

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10
Not at all certain Somewhat certain Very certain

6. What do you think the test that asked you to remember line-drawings was designed to measure?

How certain are you about your answer?

1 2 3 4 5 6 7 8 9 10
Not at all certain Somewhat certain Very certain

11. What do you think the test that required you to tap your finger as fast as you could was designed to measure?

____

____

____

____

How certain are you about your answer?

____

1	2	3	4	5	6	7	8	9	10
Not at all certain				Somewhat certain				Very Certain	

Appendix F

Instructions for Coached Brain Injury Simulators – Early Warning

Please follow these instructions for the remainder of the experiment ***without*** letting the examiner know what you have been told to do! When you finish reading these instructions, sign at the bottom indicating that you have read them and understand your task. Then, place this signed sheet back into the envelope, seal it, place an X over the seal and wait for the examiner to return. You will be asked about these instructions later on.

You are about to take some cognitive tests that examine mental abilities such as attention, memory, thinking, and reasoning skills, and your ability to think quickly. It is possible that some of the tests you will take today were designed to detect if someone is faking brain damage.

While responding to the tests, please pretend that you have experienced brain damage from a car accident involving a head-on collision. You hit your head against the windshield and were knocked out for 15 minutes. Afterwards, you felt "dazed" so you were hospitalized overnight for observation. Because the driver of the other car is at fault, you have decided to go to court to get money from the person responsible. During the next few months following the accident, the negative effects from your head injury disappear. Your lawsuit has not been settled yet, and your lawyer has told you that you may get more money if you look like you are suffering from brain damage.

As you pretend to be this car accident victim, try to respond to each test as a patient who is trying to appear brain damaged in order to get money from the lawsuit. Thus, your performance on the tests should convince the examiner as well as the people involved in deciding the outcome of your lawsuit that you are still suffering from brain damage. In order to convince these individuals, your brain damage must be *believable. If you succeed in convincing the examiner that you have suffered a brain injury, you will receive two additional experimental credits, for a total of 6 credits.*

Try to produce the most severe problems that you can *without* making it too obvious to the examiner. Major exaggerations, such as remembering absolutely nothing, are easy to detect. If the examiner does not believe that you have any problems you will not get any money for your head injury. People who have a head injury often have problems paying attention, cannot remember things as well, and do not learn things as easily as they did before their injury. They also think a little slower than they used to. Keep this in mind when taking the tests. Remember you are to try to mimic the performance of persons who are truly brain damaged.

During the time that the examiner is out of the room, you may prepare for the examination. Please wait for the examiner to return to the room. Remember to sign this sheet, place it in the envelope, and seal the envelope before the examiner returns. Again, do not let the examiner know what these instructions have told you to do. Your performance on the tests should be the only way someone could figure it out.

Thank you very much for your participation.

I have read these instructions and will do my best to follow them for the remainder of the experiment.

____

(*Signature)*

Instructions for Coached Brain Injury Simulators – Late Warning

Please follow these instructions for the remainder of the experiment *without* letting the examiner know what you have been told to do! When you finish reading these instructions, sign at the bottom indicating that you have read them and understand your task. Then, place this signed sheet back into the envelope, seal it, place an X over the seal, and wait for the examiner to return. You will be asked about these instructions later on.

You are about to take some cognitive tests that examine mental abilities such as attention, memory, thinking, and reasoning skills, and your ability to think quickly.

While responding to the tests, please pretend that you have experienced brain damage from a car accident involving a head-on collision. You hit your head against the windshield and were knocked out for 15 minutes. Afterwards, you felt "dazed" so you were hospitalized overnight for observation. Because the driver of the other car is at fault, you have decided to go to court to get money from the person responsible. During the next few months following the accident, the negative effects from your head injury disappear. Your lawsuit has not been settled yet, and your lawyer has told you that you may get more money if you look like you are suffering from brain damage.

As you pretend to be this car accident victim, try to respond to each test as a patient who is trying to appear brain damaged in order to get money from the lawsuit. Thus, your performance on the tests should convince the examiner as well as the people involved in deciding the outcome of your lawsuit that you are still suffering from brain damage. In order to convince these individuals, your brain damage must be *believable. If you succeed in convincing the examiner that you have suffered a brain injury, you will receive two additional experimental credits, for a total of 6 credits.*

Try to produce the most severe problems that you can *without* making it too obvious to the examiner. Major exaggerations, such as remembering absolutely nothing, are easy to detect. If the examiner does not believe that you have any problems you will not get any money for your head injury. People who have a head injury often have problems paying attention, cannot remember things as well, and do not learn things as easily as they did before their injury. They also think a little slower than they used to. Keep this in mind when taking the tests. Remember you are to try to mimic the performance of persons who are truly brain damaged.

During the time that the examiner is out of the room, you may prepare for the examination. Please wait for the examiner to return to the room. Remember to sign this sheet, place it in the envelope, and seal the envelope before the examiner returns. Again, do not let the examiner know what these instructions have told you to do. Your performance on the tests should be the only way someone could figure it out.

Thank you very much for your participation.

I have read these instructions and will do my best to follow them for the remainder of the experiment.

____

(*Signature)*

Instructions for Controls – Early Warning

Please follow these instructions for the remainder of the experiment *without* letting the examiner know what you have been told to do! When you finish reading these instructions, sign at the bottom indicating that you have read them and understand your task. Then, place this signed sheet back into the envelope, seal it, place an X over the seal and wait for the examiner to return. You will be asked about these instructions later on.

You are about to take some cognitive tests that examine mental abilities such as attention, memory, thinking and reasoning skills, and your ability to think quickly. It is possible that some of the tests you will take today were designed to detect if someone is faking brain damage.

Your task is to perform to the best of your ability, answering all questions in as honest a fashion as you can.

While the examiner is out of the room, remember to sign this sheet, place it in the envelope, seal it, and place an X over the seal of the envelope before the examiner returns. Please wait for the examiner to return to the room.

Again, do not let the examiner know what these instructions have told you to do. Your performance on the tests should be the only way someone could figure it out.

Thank you very much for your participation.

I have read these instructions and will do my best to follow them for the remainder of the experiment.

____

(*Signature*

Instructions for Controls – Late Warning

Please follow these instructions for the remainder of the experiment *without* letting the examiner know what you have been told to do! When you finish reading these instructions, sign at the bottom indicating that you have read them and understand your task. Then, place this signed sheet back into the envelope, seal it, place an X over the seal and wait for the examiner to return. You will be asked about these instructions later on.

You are about to take some cognitive tests that examine mental abilities such as attention, memory, thinking and reasoning skills, and your ability to think quickly.
Your task is to perform to the best of your ability, answering all questions in as honest a fashion as you can.

While the examiner is out of the room, remember to sign this sheet, place it in the envelope, seal it, and place an X over the seal of the envelope before the examiner returns. Please wait for the examiner to return to the room.

Again, do not let the examiner know what these instructions have told you to do. Your performance on the tests should be the only way someone could figure it out.

Thank you very much for your participation.

I have read these instructions and will do my best to follow them for the remainder of the experiment.

____
(*Signature)*