THE VALIDITY OF THE LETTER MEMORY TEST AS A MEASURE OF
MEMORY MALINGERING: ROBUSTNESS TO COACHING

A dissertation presented to
the faculty of
the College of Arts and Sciences of Ohio University

In partial fulfillment
of the requirements for the degree
of Doctor of Philosophy

Becca L. Greub
November 2004
This dissertation entitled

THE VALIDITY OF THE LETTER MEMORY TEST AS A MEASURE OF
MEMORY MALINGERING: ROBUSTNESS TO COACHING

BY

BECCA L. GREUB

has been approved for

the Department of Clinical Psychology

and the College of Arts and Sciences by

Julie Suhr

Associate Professor of Clinical Psychology

Leslie A. Flemming

Dean, College of Arts and Sciences
The present study examined the utility of the Letter Memory Test (LMT) in the identification of malingers of cognitive deficits. The LMT is a computerized forced-choice test that includes two face validity manipulations: increase in stimulus length and increase of response choices. Performance on the LMT was compared in healthy controls and in participants with head injury, some of whom were asked to perform with best effort and others asked to simulate cognitive impairment following head injury. As an additional manipulation, those asked to simulate head injury were either warned about the potential for malingering detection or not. The LMT was administered in the context of a clinically relevant battery of neuropsychological assessments, using a counterbalanced design regarding early or late administration of the LMT and the 15-Item Test (FIT).

A cutoff of 9 correct on the FIT and 93% correct on the LMT were used. As expected, the LMT was found to be insensitive to head injury, but sensitive to malingering and robust to detection warning. When comparing the two simulator groups (warned and naïve) with the two best effort groups (head injury and healthy), the LMT yielded 76% sensitivity, 96% specificity, 95% positive predictive accuracy and 81% negative predictive accuracy. The FIT yielded 17% sensitivity, 100% specificity, 100% positive predictive accuracy, and 56% negative predictive accuracy. Additionally, the LMT correctly identified all of the malingerers that were identified by the FIT, but the FIT missed 32 of the malingerers that were identified by the LMT. Detrimental order
effects were found for the FIT but not the LMT. The two face-validity manipulations of the LMT were also examined. For the two best effort groups and the warned simulator malingerers, performance across both target stimuli length and number of responses remained stable, while performance by the naïve group worsened as stimulus length and number of response choices increased.

These results replicate earlier findings that indicate the LMT is an adequate measure of malingering of cognitive deficits and superior to the FIT. The results also extend earlier findings by demonstrating relative robustness of the LMT to coaching.

Approved: Julie Suhr

Associate Professor of Clinical Psychology
Dedication

To my parents, Louis and Barbara, for your continued support through all the years of my life. I will be honored to be able to share the title “Dr. Greub” with my father.
Acknowledgements

I wish to thank several individuals for their contributions and support during this research. First and foremost, I would like to thank the chair of my dissertation committee, Dr. Julie Suhr, for her patient guidance and support during this long project. As my advisor, she is a mentor and model of the type of professional I hope to become. I am honored to have a prior publication with Dr. John R. McNamara and greatly appreciate his participation as a committee member to review this manuscript. In addition, I am grateful for the time and contributions of committee members of Dr. Ben Ogles, Dr. Timothy Anderson, and Dr. Tracy Leinbaugh for their review and critique of this manuscript.
Table of Contents

Abstract 4

Dedication 6

Acknowledgements 7

List of Tables 11

List of Figures 12

Literature Review 13

Indices of Accuracy 15

The Base-Rate Factor 16

Sample Characteristics And Considerations 18

Coaching 21

Methods of Memory Malingering Detection 25

Cutoff Score Method 25

Patterns of Performance Method 28

Tests Developed Specifically For Malingering Detection 33

Forced-Choice Method 37

Hypotheses of the Present Study 48

Method 50

Participants 50

Procedure 51

Neuropsychological Measures 53
Table of Contents (continued)

Results  59

   Performance on the LMT  59
   Performance on the 15-Item Test  61
   Order effects of the LMT and 15-Item Test  61
   Hypothesis 1: Sensitivity of the LMT to Malingering  62
   Hypothesis 2: Specificity of the LMT to Malingering  63
   Hypothesis 3: Robustness of LMT to Coaching  64
   Hypothesis 4: Comparison of LMT to the 15-Item Test  65
   Hypothesis 5: Patterns of Performance on the LMT  67
   Non-Malingering Measures  73

Discussion  75

   LMT Versus 15-Item Test  76
   Effects of Warning  77
   Limitations of the Present Study  79

Future Research Directions  80

   Research Paradigms  80
   Definition of Malingering  81
   The Daubert Standard  83
   Reaction Time  84
   Coaching and Warning  85
   Multiple Measures of Malingering  86

References  89
<table>
<thead>
<tr>
<th>Appendix A: Mass Screening Questionnaire</th>
<th>102</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix B: Consent Form</td>
<td>103</td>
</tr>
<tr>
<td>Appendix C: Instructions to Participants</td>
<td>104</td>
</tr>
<tr>
<td>Appendix D: Compliance Questionnaire</td>
<td>105</td>
</tr>
</tbody>
</table>
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Forced Choice Measures of Malingering</td>
<td>44</td>
</tr>
<tr>
<td>2. Timing of Measures</td>
<td>54</td>
</tr>
<tr>
<td>3. Demographics By Group</td>
<td>60</td>
</tr>
<tr>
<td>4. Total LMT Scores By Group</td>
<td>60</td>
</tr>
<tr>
<td>5. Total 15-Item Test Scores By Group</td>
<td>61</td>
</tr>
<tr>
<td>6. Simulator Performance on LMT and 15-Item Test Based on Administration Status</td>
<td>63</td>
</tr>
<tr>
<td>7. Group Comparisons and Indices of Accuracy Using 93% Cutoff on LMT</td>
<td>65</td>
</tr>
<tr>
<td>8. Group Comparisons and Indices of Accuracy Using 9 Correct Cutoff on 15-Item Test</td>
<td>67</td>
</tr>
<tr>
<td>9. Descriptive Statistics For Stimulus Levels By Group</td>
<td>68</td>
</tr>
<tr>
<td>10. Within Subject Effects By Group For Stimulus</td>
<td>69</td>
</tr>
<tr>
<td>11. Descriptive Statistics For Response Levels By Group</td>
<td>71</td>
</tr>
<tr>
<td>12. Within Subject Effects By Group For Choice</td>
<td>72</td>
</tr>
<tr>
<td>13. Group Performances On Non-Malingering Measures</td>
<td>74</td>
</tr>
</tbody>
</table>
List of Figures

Figure

1. Indices of Accuracy 15
2. LMT Percent Correct on Stimulus Length By Group 72
3. LMT Percent Correct on Response Options By Group 73
The Validity Of The Letter Memory Test As A Measure Of Memory Malingering: Robustness To Coaching

The Diagnostic and Statistical Manual of Mental Disorders 4th Edition (DSM-IV), defines malingering as "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" (APA, 1994). In the field of neuropsychology, individuals may feign neuropsychological deficits, or mangle, by consciously manipulating their performance on cognitive tests in an effort to obtain external incentives. Such incentives may include significant legal or monetary compensation such as disability claims, avoidance of criminal conviction or incarceration, or avoidance of unwanted duty. In other words, malingerers are purposefully assessing their own performance and deliberately attempting to perform significantly below their abilities for the purpose of secondary gain.

Although the DSM-IV addresses malingering of any disorder, researchers in the field of clinical neuropsychology have identified problems with the definition and criteria, particularly when applied to malingering of cognitive impairment. Slick, Sherman, and Iverson (1999) view the malingering definition by the DSM-IV as too vague and its classification as an issue that may need clinical attention and not a legitimate mental disorder as inadequate. They proposed a more specific and objective definition, which included intentional exaggeration of symptoms or fabrication of deficits. Also included is the intent to achieve secondary gains such as financial awards, or the avoidance of duties that one is legally required to perform. Their proposed criteria include the presence of substantial external incentive and evidence for negative response bias (i.e., poor performance on tests specifically designed to detect malingering of
cognitive impairment) not better accounted for by psychiatric, neurological, or developmental factors. Slick et al. further propose the diagnostic qualifiers of “definite,” “probable,” and “possible.” By their definition, external incentives, along with evidence for a definite negative response bias (such as performing below chance on an instrument designed to detect cognitive malingering) indicate definite malingering. Probable malingering includes external incentive as well as two or more of the following obtained by neuropsychological malingering measures: probable response bias, discrepancy between test data and known patterns of brain functioning, discrepancy between test data and observed behavior, discrepancy between test data and reliable collateral reports, and discrepancy between test data and documented background history. Therefore, the criteria for malingering rely heavily on the respondent’s performance on neuropsychological testing, specifically performance on instruments designed to detect malingering, and it would be judicious to further develop this vein of research.

It is beneficial to accurately detect malingering. Positively identifying malingerers reduces both the inefficiency of services and unwarranted legal or monetary compensation. Yet false identification of malingerers may result in the withholding of services for those in need. The present study was designed to provide further evidence for the validity of a recently developed memory malingering instrument. First, psychometric issues necessary to consider in malingering detection research are examined, including the indices of accurate detection in the context of malingering base-rates, population and sample characteristics and considerations, and coaching issues. The wide range of memory malingering detection instruments will be reviewed in the context of these psychometric concerns.
Indices of Accuracy

Rosenfeld, Sands, and Van Gorp (2000) review four indices of accuracy in the identification of malingering (see Figure 1). Sensitivity and specificity are viewed in the context of prediction and are properties of the neuropsychological test being utilized. Sensitivity is the proportion of correctly predicted malingerers, while specificity is the proportion of correctly predicted non-malingerers.

![Figure 1](image_url)

### Indices of Accuracy

<table>
<thead>
<tr>
<th>Actual</th>
<th>Malingering</th>
<th>Non-malingering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malingering</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Not Malingering</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

**Index**

- **Sensitivity**: $A/(A + C)$
- **Specificity**: $D/(B + D)$
- **PPA**: $A/(A + B)$
- **NPA**: $D/(C + D)$

Note: PPA = positive predictive accuracy, NPA = negative predictive accuracy

In other words, sensitivity reflects malingerers who are correctly identified as malingerers by the test, while specificity reflects those correctly identified as honest
respondents by the test. Predictive accuracy views malingerers and non-malingers in the context of actual condition and is a property of the both the test and the population being examined. Positive predictive accuracy is the percentage of people correctly classified as malingering when the test predicts malingering, while negative predictive accuracy is the percent of people correctly classified as non-malingering when the test predicts non-malingering.

Rosenfeld and his colleagues (2000) and Rogers (1997) contend positive predictive accuracy is the most clinically relevant of the indices. However, the indices are interrelated. Positively identifying malingerers may result in an increase in false negatives. As noted previously, positively identifying malingerers reduces both the inefficiency of services and the unwarranted legal or monetary compensation. Yet false identification of malingerers may result in the withholding of services for those in need. The challenge is then to satisfy the consideration of all four of the indices, thereby honing the accuracy of identification. Furthermore, as opposed to identifying tests that can be utilized in all settings and populations, it may be more effective and practical to explore instruments that maximize positive predictive accuracy in specific settings and contexts.

The Base-Rate Factor

A factor to be considered that supports setting- and context-specific research is the problematic variable base-rate of malingering. In standard clinical neuropsychological settings, estimated base rates cited by review studies of malingering range from 15% to up to 60% (Greiffenstein, Baker, & Gola, 1994; Rogers, Harrell & Liff, 1993). In forensic settings, rates are similar, with estimates of malingering varying from 17% to 64% (Binder & Willis, 1991; Heaton, Smith, Lehman & Vogt, 1978; Rogers, Salekin, Sewell,
Goldstein, & Leonard, 1998; Youngjohn, 1991). One reason for such different percentages is the varied definitions of malingering used by different researchers. The most consistent aspects of malingering definition in these studies are sufficiently low performance on neuropsychological tests (i.e., performing below chance or performing significantly worse than an impaired population that was not involved in litigation) and evidence of incentive to malinger (workman’s compensation, litigation). Provided that there has been a significant increase in benefits and compensation of personal injury cases over the past two decades, the malingering rate is likely to have been maintained or increased (Bates & Groussman, 2000). Perhaps since the development of malingering criteria proposed by Slick et al. (1999) discussed above, future research on base rates of malingering may be more unified. Another factor that may contribute to the varied rate of malingering is the types of cases that are referred to a particular setting, as well as the general clinic setting (inpatient, outpatient, private practice).

Although some of the variation in these base-rates may appear to be minor, a variation in base-rates can significantly impact predictive accuracy and result in an increase in diagnostic errors. For example, Rosenfeld et al. (2000) manipulated the data of Mittenberg, Azrin, Millsaps, and Heilbronner (1993) after Mittenberg et al. set a 50% base-rate to determine an optimal score for delineating malingers from non-malingerers on a measure. Rosenfeld et al. found that when sensitivity and specificity were constant, .77 and .90, respectively, even a 5% change in base-rates significantly influenced predictive accuracy.
Sample Characteristics and Considerations

Another methodological consideration in malingering research is the research sample. Ideally, research on malingering includes a clinical sample of individuals known to be malingering. Obtaining such a sample would provide the most accurate reflection of a true clinical situation. Norms generated on a known clinical sample would be more generalizable to malingerers not included in the research sample, relative to other samples such as a simulated malingering population. For example, examining the performance of individuals who have experienced a blow to the head in a specific area and who meet diagnostic criteria for malingering would more accurately reflect performance of other unknown malingerers with a similar injury. By using a sample of non-injured individuals who are instructed to malinger, generalizability and accuracy are potentially reduced, due to factors such as inaccurate malingering performance portrayal and lack of realistic motivation to malinger convincingly. However, it is difficult to know who is actually malingering in clinical samples; individuals who are malingering are not apt to readily reveal this information. Thus, deciding who in a clinical sample is malingering requires some a priori rules, none of which have been agreed upon in the existing literature (Suhr, Tranel, Wefel, & Barrash, 1997). For example, assuming that all persons in litigation are malingering is likely to be inaccurate. Suhr et al. found that litigation status alone was not a satisfactory predictor of malingering or exaggeration of deficits. Also, many studies create confounds when testing the accuracy of new malingering tests by using similar tests to define malingering in their samples. Slick et al. proposed more standard and comprehensive diagnostic criteria, discussed above.
However, using performance on malingering tests as both definition of malingering and as a dependent variable may continue to present as a confound to research in this area.

Since it is difficult to determine a clinical malingering sample with certainty, many researchers have utilized individuals who simulate or role-play head injury performance in studies of malingering detection. A simulator sample can mimic the purposeful poor performance on cognitive tests in an effort to appear convincing and can easily be identified as malingering (as instructed by the researcher). However, generalizability is sacrificed for a known malingering population. Debates continue to exist not only over whether or not the simulated malingers have enough motivation to attempt to mangle convincingly, but also in how well simulated malingers can accurately feign cognitive impairment without having real world experience of head injury. A number of studies have explored possible means to increase motivation of undergraduate student simulators. One manipulation has been to provide financial incentives to the participants to provide motivation to feign neurological deficits as realistically as possible (Rogers, 1997; Rogers et al., 1993). This incentive is an effort to simulate the similar (although much more significant) financial gain obtained by real world malingers and thus increase the generalization of findings. One means of providing a financial incentive is via a direct monetary reward to participants ranging from $20-$30 for motivation to malinger (Heaton et al., 1978; Inman & Berry, 2002; Orey, Cragar & Berry, 2000; Vickery, 2001). However, due to ethical considerations, all malingerers, regardless of motivation, were either provided with the bonus credits or the money.
A more practical means to provide financial incentive is to incorporate a lottery system. Iverson and Franzen (1996) and Dunn, Shear, Howe, and Ris (2003) incorporated a $100 lottery system to the simulated malingering condition. Iverson and Franzen offered five $100 incentives to 135 participants, while Dunn et al. offered one $100 incentive to 60 participants. A monetary lottery may be a practical means to provide monetary incentive; however, participants may be aware of low odds or low pay-off compared to the thousands of dollars sought in litigation and be unaffected by the motivator effort.

Although several recent researchers have implemented financial reward procedures in their studies, very few have evaluated the effect of the incentive implemented; those researchers that did have found that the motivator did not significantly impact cognitive performance (Bernard, 1990; Orey et al., 2000).

The second relevant issue is whether or not simulators can malinger cognitive deficits due to head injury in a believable manner. One method that attempts to compensate for this factor is utilizing a sample of healthy participants familiar with the effects of head trauma (Strauss, Slick, Levy-Bencheton, Hunter, MacDonald, & Hultsch, 2002) such as medical, rehabilitation, or legal professionals. Another method is to include individuals with previous head injury experience and thus have some real-world experience with the consequences of head injury such as head trauma accompanied by a loss of consciousness (Inman & Berry, 2002; Orey et al., 2000).

In order to compensate for some of these limitations to using a simulator sample, the present study included clinically relevant contextual factors, in an effort to replicate actual clinical experience. The study utilized participants with a history of head trauma
accompanied by loss of consciousness. In addition, they were administered a comprehensive battery of neurological tests, similar to a true clinical experience, rather than isolated malingering tests.

Coaching

An additional factor that is important to consider in research on malingering detection is that of coaching of the participant. Coaching involves providing specific information on how to counteract malingering checks, or providing strategies on how best to simulate head injury. Sophisticated individuals may be able to mangle in a convincing manner. Wetter and Corrigan (1995) found a substantial portion (48%) of lawyers took the responsibility to prepare a client for psychological testing, which included a discussion of validity measures within those instruments. Providing information to a respondent regarding how to malinger convincingly has significantly affected the ability of neuropsychological measures to identify malingerers. Lees-Haley (1997) reported that attorneys are influencing data reviewed by psychological experts via suggestions on how to respond to psychological tests, suggestions of what to emphasize during psychological evaluations, and advising patients what information to disclose and what information to withhold.

Research on coaching reflects various levels of disclosure. Some researchers provide coached malingerers with specific information on how to appear as a bona fide head injured respondent. Storm and Graham (2000) found that coached simulators who were provided tips on how to avoid detection successfully evaded detection by the Minnesota Multiphasic Personality Inventory – 2 (MMPI-2) F scale, yet uncoached malingerers who were merely asked to fake deficits in a believable manner were
identified. Lamb, Berry, Wetter, and Baer (1994) used college student simulators and examined the effect of coached information on the MMPI-2. Experimental groups were provided with head trauma information, test validity scale information, or both. A control group was also included. The researchers found that information on validity scales decreased scores on both the validity and clinical scales while information on head trauma inflated the F scale which is aimed at measuring the over-reporting of symptoms or “faking bad.”

DiCarlo, Gfeller, and Oliveri (2000) administered the Category Test to coached simulators who were instructed on how to present cognitive impairment and given strategies on how to avoid detection, uncoached simulators who were merely asked to feign impairment, optimal-performance controls who were instructed to perform to the best of their ability, and a control group of traumatic brain injury patients who were also instructed to perform to their best ability. The researchers found that significantly more coached simulators were misclassified as performing genuinely than the uncoached simulators.

Similarly, Rose et al. (1998) administered four memory malingering measures (the Portland Digit Recognition Test – Computer, Rey’s Dot Counting Test, The 21-Item Test, and the Nonverbal Forced Choice Test) to a sample of healthy college students asked to simulate malingering, healthy students asked to simulators who were provided information on head injury, non-litigating head injured patients, and healthy controls. The researchers found that coached simulators who were provided with specific information on how to present with head injury deficits were better able than uncoached simulators in avoiding detection. Combined, the four tests correctly classified 100% of the healthy
controls, 47% of the uncoached simulators but only 29% of the coached simulators. The researchers further noted that on the indices of the Portland Digit Recognition Test total score, Nonverbal Forced Choice Test-total correct and consistency ratio, and the 21-Item Test forced-choice recognition score, the coached simulators performed significantly better than the uncoached simulators, indicating less impairment and suggesting these tests were not robust to coaching.

Recent research shows that even very subtle coaching (i.e., warning a participant that a neuropsychological test battery may include attempts to detect malingering) can alter performance. Suhr and Gunstad (2000) found that sensitivity of a forced choice memory malingering test in identifying simulators was reduced when head injury simulators were warned that there might be an attempt to detect malingering on the tests administered, without providing any further coaching about how to avoid detection. Furthermore, the researchers found that warned simulators performed significantly better than naïve simulators. In a similar study, Gunstad and Suhr (2001) found that warned simulators were better at avoiding identification on a forced choice memory malingering test than naïve simulators or simulators informed of typical brain injury symptoms with no warning about detecting malingering.

Some tests may be more vulnerable to coaching than others. Feldstein, Durham, Keller, Klebe, and Davis (2000) found that malingers with no or low coaching performed significantly worse than those given a higher level of coaching on a non-declarative dots categorization task and that individuals with higher levels of coaching were misclassified as impaired controls on an easy (less distorted) version of the dots categorization test. This information may indicate that the face-validity of a task may be a
significant factor in malingering and coaching. Suhr and Gunstad (2000) found that violations of typical yet less obvious task performance patterns (i.e., differences between basic attention and memory, failure to recognize words previously recalled three times in learning trials) were more robust to malingering than a face valid forced-choice malingering task. This research indicates a need for malingering measures that are robust to the various levels of coaching and incorporate face-validity manipulations to complicate the task of the malingerer.

Ethical dilemmas have become salient in the study of malingering. Ben-Porath (1994) suggested that, by researching malingering, the information may be used for the unethical purpose of teaching attorneys and individuals how to malinger convincingly. Berry, Lamb, Wetter, Baer, & Widiger (1994) described a dilemma of satisfying APA ethical guidelines of protecting the integrity of psychological tests and understanding the impact of coaching on such tests. Berry and his colleagues purport that a resolution of this issue is to reveal the vulnerability of tests to malingering, yet not reveal details of successful coaching strategies. Furthermore, Youngjohn, Lees-Haley, and Binder (1999) found that merely warning a respondent that malingering checks are in place reduces the validity of forensic evaluations. Since even a minor warning about the potential for malingering detection appears to alter performance (Gunstad & Suhr, 2001; Suhr & Gunstad, 2000), research need not provide damaging information to the respondent to study malingering. The present study avoided such ethical dilemmas by focusing on the impact of test ability to detect malingering by using a simple warning to the respondent of detection attempts and measuring results and using a simulator sample, rather than a
patient sample for which the clinical data may be required for medical care or legal
decision making. This research strategy satisfies issues raised by Ben-Porath.

Methods of Memory Malingering Detection

Three procedures involving neuropsychological tests are popular in the
identification of malingerers; 1) cutoff scores of standard neuropsychological tests that
differentiate malingerers from honest but brain damaged responders, 2) the qualitative
method of identifying meaningful learning and memory patterns breached by
malingerers, and 3) the development of measures specifically designed to detect
malingering, particularly the forced choice method.

Cutoff Score Method

One approach to the detection of malingering has been to identify cutoff scores on
standard psychological tests that separate malingerers from non-malingerers. The premise
of this method dictates that the performance of malingerers is consistently and
sufficiently worse than bona fide impaired individuals. A number of tests have been
utilized in this method.

The California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober,
1987) is a 16–item serial learning task. The items of the shopping lists are associated with
days of the week. This test measures verbal memory as well as conceptual ability and can
reveal learning strategies. The test provides two recall trials, one with a category
provided as a memory cue. This test yields a number of scores including total immediate
recall, rate of learning, recall consistency, serial position, free and cued recall, recall and
recognition, and intrusion and repetition errors.
Trueblood and Schmidt (1993) found that in their sample of mild head injured patients involved in litigation, those litigants that scored below chance on a forced choice task, also scored worse on the CVLT recognition task compared to litigants who did not perform poorly on the forced choice task.

Millis, Putnam, Adams, and Ricker (1995) explored the utility of the CVLT in the identification of non-litigating patients with bona fide moderate to severe brain injury from litigating mild head injured patients who performed at or below chance on a forced-choice memory test, indicating incomplete effort. The researchers focused on four patterns of performance of the CVLT, including Trials 1-5 of List A, recognition discriminability, recognition hits, and long-delay cued recall. The CVLT Total Trials 1-5 List A yielded a correct classification rate of 83% with 74% sensitivity and 91% specificity. Setting a lower specificity limit at 90%, recognition discriminability yielded a sensitivity of 96%, specificity of 91%, and overall correct classification rate of 93%. Individual cutoff scores for long-delay cued recall obtained an 87% correct classification rate, with 83% sensitivity and 91% specificity.

The Recognition Memory Test (RMT; Warrington, 1984) is a test that assesses the ability of a respondent to recognize a visual stimulus (word or a face) when presented with a target stimulus and a foil. Iverson and Binder (2000) reviewed normative RMT data for a variety of patient groups, litigants with head injury, and known malingers. This data can provide cutoff scores for different groups of respondents. The comparability of the performance of a respondent compared to the known performance of the variety of patient groups can prove useful in the identification of malingers. The researchers found that simulated analog malingers performed more poorly than patients
with brain injury, patients following brain surgery, and even patients with Alzheimer’s disease. The performance of the malingerers was most similar to the performance of litigants claiming head injury.

Iverson and Franzen (1996) determined cutoff scores for a number of memory tasks, including the digit span subtest from the Wechsler Memory Adult Intelligence Scale – Revised, the 21-Item test (discussed below), and a logical memory subtest of the Wechsler Memory Scale – Revised. Their sample consisted of 20 undergraduates and 20 psychiatric inpatients, some of who were asked to malinger while the others were asked to perform to their ability. Twenty memory impaired patients also participated. These patients were identified as memory impaired based on their performance on The Luria Nebraska Neuropsychological Battery or the Wechsler Memory Scale – Revised. The researchers found that a cutoff score of less than 4 digits on digit span forward resulted in correct classification of 60% of the malingerers, 100% of the normal effort group, and 100% of the memory impaired participants, with an overall rate of 70% correct classifications and zero false positives. On digit span backwards, a cutoff score of less than three yielded a 57.5% correct classification of malingerers, 97.5% correct classification of the normal effort group, and 95% correct classification of memory impaired participants, with an overall rate of 81% of correct classification and a 3% false positive rate.

Bernard, McGrath, and Houston (1996) review the advantages of using the discriminant function analysis (cutoff) method. It produces known specificity and sensitivity rates, it can be applied to standardized tests, and discriminant functions can be derived for various patient and non-patient groups. However, a number of limitations are
relevant to using the relatively simple cutoff method of identification. A lack of comprehensive normative data results in research suggesting multiple cutoff scores for the same test. A major limitation of the cutoff method stems from the blurry delineation of the performances of malingerers and impaired individuals. Although research on specific injury type (such as head injury victims), indicates that malingerers often perform more poorly than individuals with a head injury, the range of performance in the two groups often overlap and inaccurate identification is a reflection of the sometimes vast overlap in performance in both the experimental and clinical setting.

Another criticism is the lack of replication of findings in individual studies. The resulting cutoff scores from discriminant function analysis may not be accurate when applied to another sample (Iverson & Franzen, 1996), indicating a need for independent replication and cross validation of discriminant formulas in other samples (Franzen et al., 1990).

Regardless of the criticisms of the cutoff method, any one assessment should not be used alone in the diagnosis of malingering. The cutoff method can be useful in looking for patterns of performance of malingerers by examining multiple cutoffs within an assessment (such as subscales) or even cutoffs of a number of assessments. Identifying patterns of performance that are more subtle and complex in assessments is an additional method of malingering detection.

*Patterns of Performance Method*

In addition to the quantitative method of using cutoff scores, recent research on malingering includes examining meaningful patterns of response on neuropsychological tests. The premise of this method is that these patterns of responding include subtle
features typically unattended to by the malingerer. For example, the Wechsler Memory Scale – Revised (WMS-R; Wechsler, 1987) consists of 9 subtests, which collectively measure orientation, attention, and immediate and delayed recall of both verbal and visual information. Mittenberg et al. (1993) examined the performance of a group of non-litigating head injured participants referred for a neuropsychological evaluation and a matched healthy control group who were asked to simulate malingering on the WMS-R. The researchers found that the simulated malingerers obtained reliably lower scores than the head injury group on the Attention/Concentration index, Mental Control, Figural Memory, Digit Span, Visual Memory Span, Visual Pair Associates, and Verbal Paired Associates. Also, the simulated malingerers obtained significantly higher General Memory Index scores than Attention/Concentration scores while the head injury group yielded significantly higher Attention/Concentration index scores than General Memory Scores. The researchers found that with a discriminant function analysis with the General Memory score minus Attention/Concentration score, the test yielded an 83.3% correct classification rate with a false positive rate of 10.3% and a false negative rate of 23.1%.

A pattern may be as simple as combining scores and applying a cutoff or more complex such as learning patterns. This can be illustrated, for example, in examining a malingerer’s performance on the U shaped learning curve associated with memory of serial items. The primacy effect is a well-established list learning pattern, indicating that items at the beginning of an item list are more likely to get recalled, due to more rehearsal and higher likelihood of being processed into long term memory. Also, items at the end of the list are still being processed by short-term memory and readily able to be recalled
(the recency effect). Thus, beginning items and the ending items are more readily recalled than items in the middle of a word list.

Hall and Bornstein (1991) found the serial position curve intact for minor closed head injury patients and matched healthy controls. Bernard (1991) explored the serial position curve via the Auditory Verbal Learning Test with a head injury malingering group, a group of closed head injury patients, and a healthy control group. Malingers demonstrated a disruption of the primacy effect, while the head injury patients demonstrated a disruption of the recency effect. In a second study, Bernard, Houston, and Natoli (1993) found that simulated malingerers showed no disruption of the serial position curve.

Suhr et al. (1997) compared head injury patients who were anticipated to be probable malingerers with mild head injured patients in litigation not anticipated to be malingerers, mild to moderately head injured patients not involved in litigation, severe head injured patients not involved in litigation, and psychiatric patients with no history of head injury. Results showed a normal serial position curve for all groups except those identified as probable malingerers. This group demonstrated a suppressed primacy effect. Even closed head injury patients demonstrated a primacy effect. Suhr (2002) examined serial position, warning, and learning and delayed recall. She found that both naïve and warned simulated malingerers demonstrated a suppression of the primacy effect, while healthy and head injury participants demonstrated normal serial position curves.

Other factors such as memory inconsistencies, recognition, recall, and learning span are also useful in assessment, identifying meaningful patterns of malingering. One of the more useful tests available in this area is the Auditory Verbal Learning Test.
(AVLT; Rey, 1964), which allows the researcher to incorporate memory, learning, recall, recognition, memory inconsistencies, and learning span. In contrast to specific cutoff scores, the AVLT produces qualitative information. For example, Suhr et al. (1997) noted that patients with brain injury typically recognize words that they have been able to freely recall during the majority of learning trials, while probable malingerers frequently showed no recall or recognition of the words despite prior learning performance of the words. Based on these preliminary findings, Barrash, Suhr, and Manzel (2004) developed a malingering index (AVLT-Exaggeration Index; AVLTX) based on several qualitative aspects of AVLT performance including exceedingly poor learning, lack of primacy effect, worsening recall, worsening recognition, learned words not recognized, recalled words not recognized, and exceedingly poor recognition. Barrash et al. examined the performance of probable patient malingerers who presented to a clinic with complaints of memory impairment with minimal brain injury and issues of secondary gain. Also included in the study was a group of patients with bona fide brain damage and a group of psychiatric patients. The researchers found that the cutoff of 3 on the AVLTX yielded 71% sensitivity with the probable patient malingerers, 94% specificity with the brain damage group, 92% specificity with the psychiatric patients, and a positive predictive power of 86%. In a cross-validation study, the researchers included the same three sample groups (this implies that it was not independent sample, reword to reflect new samples but the same diagnoses), however, financial issues were more prevalent for each of the three groups. Despite the presence of financial issues, the AVLTX yielded high specificity with the brain damage group, as only 2.5% false negatives were found when a cutoff of 3 was used. The specificity of the psychiatric patients also remained high.
Sensitivity for probable malingerers was 57%. When the researchers compared the utility of the AVLTX with two other tests of malingering, sensitivity was comparable or superior, as was specificity for both the brain damaged group and psychiatric disorder group.

Suhr, Gunstad, Greub, and Barrash (in press) further examined the utility of the AVLTX, specifically, robustness to warning. Three groups of undergraduate students were administered the AVLTX. The experimental groups included naïve simulators and warned simulators. A best effort group was also included in the study. All groups were administered the expanded version of the AVLT in a battery of tests. However, the warned simulators were given information that indicated on of the tests in the battery was designed to identify malingering. When comparing the two simulator groups with the best effort group, the cutoff of 3 or greater yielded 23% sensitivity and 100% specificity. When comparing the naïve and warned simulators to each other, the test correctly identified 27% of the warned simulators and 19% of the naïve simulators. There was no reliable difference in the detection rates of the warned and naïve simulators, indicating robustness to warning. In an extension of the study, the researchers administered the AVLTX to student simulators with a history of mild head injury with loss of consciousness, some of whom were warned about the possibility of malingering detection. Sensitivity and specificity remained high, and the index was still robust to coaching.

The patterns of performance method for malingering detection, although more complex and cumbersome than individual cutoff scores, may indicate that a malingerer is busy attending to purposefully choosing the wrong answer, while more complex or more
subtle patterns of responding escape the malingering or may be more difficult to malingering convincingly. It is difficult, however, to be certain that false positive identification of malingeringers is minimized. Furthermore, replication is necessary as is the cross validation of discriminative functions due to the loss of predictive power during cross validation. More research is needed on various patterns of performance before they can be used with confidence.

*Tests Developed Specifically For Malingering Detection*

Rather than looking for scores on existing neuropsychological batteries that discriminate malingeringers from non-malingeringers, some researchers have developed instruments whose sole purpose is to measure malingering. A few such tests are based on the floor effect. These tasks are extremely easy, but malingeringers may be fooled into thinking they are harder than they seem and thus perform worse than even brain-damaged individuals. Others are based on binomial probability principles and are commonly called forced choice tests.

*Floor Effect*

The 15-item test (Rey Memory Test, FIT, Rey: 1964) presents 15 serial items to the participant who is asked to memorize and recall them. The items are presented in five rows with three characters per line: A, B, C; 1, 2, 3; a, b, c; a circle, a square, a triangle; and Roman numerals I, II, III. Most brain damage patients perform the task easily. However, the instructions make the task purposefully appear more difficult, which is intended to mislead the malingeringer into thinking the task is more difficult (Lezak, 1995). Lezak identified 9 as a sufficient cutoff score to differentiate malingeringers with honest responders. Inman and Berry (2002) used this cutoff score on a sample of non-litigating
head injured participants who were either asked to malingering or respond honestly. Also included in the study was a group of non-head injured normal effort group and non-head injured group asked to malingering cognitive deficits. The researchers found that a cutoff score of less than 9 correct resulted in 100% specificity and 2% sensitivity with an overall hit rate of 53%.

Vickery, Berry, Inman, Harris, and Orey (2001) meta-analyzed 16 studies that used the 9 correct cutoff on the FIT. They found that when comparing neurological patients responding honestly and healthy respondents asked to malingering, the FIT yielded an average of 47.7% sensitivity, 83.3% specificity, and an average hit rate of 65.1%. When comparing neurological patients responding honestly and neurological patients malingering, the test yielded an average of 50.6% sensitivity, 96.7% specificity, and an average hit rate of 85.5%. When comparing healthy participants responding honestly and healthy respondents asked to malingering, the test yielded an average 8% sensitivity, 100% specificity, and an average hit rate of 56.4%. However, when the researchers compared the FIT to other malingering measures, they found the FIT to be clearly weaker.

Lee, Loring, and Martin (1992) administered the FIT to a group of non-litigating neurological inpatients with temporal lobe epilepsy who also demonstrated memory impairment on a standardized memory test. Also included in the study was a group of non-litigating outpatients with neurological disorders and a group of litigating outpatients with neurological disorders. Litigating outpatients performed significantly worse than the other two groups. Of the inpatient group, 14% scored 9 or below while only 4% scored 7 or below. Of the non-litigating outpatients, 20% scored 9 or below while only 5% scored 7 or below. However, of the litigating outpatients, 43.75% scored 9 or below while
37.5% scored below 7. They recommended adjusting the cutoff score to 7 in effort to increase specificity.

Guilmette, Hart, Giuliano, and Leininger (1994) found 7 to be an overly sensitive cutoff score to bona fide memory impairment and somewhat insensitive to malingering. Of the honest respondents, 5% were misclassified, 40% of non-litigating patients with brain damage, and 20% of psychiatric inpatients were classified as malingerers. Millis and Kler (1995) found a cutoff score of 7 to have a low sensitivity (about 50%) of pre-identified malingerers yet no misidentification of brain-injured subjects were identified. Interestingly, Arnett, Hemmeke, and Schwartz, (1995) found that bona fide cerebral damaged respondents can achieve two or more correct rows (6 correct) in the FIT. This lends evidence to suspect significantly less damaged respondents achieving a score below 6 correct as malingerers.

Spreen and Strauss (1998) reported the reliability information is not yet available for this test. The FIT is relative quick and easy to administer and is more practical than more time consuming tests. However, Iverson and Franzen (1996) recommend that this test be given at the beginning of a battery containing more difficult tests before the respondent understands that blatant exaggeration may be easily detected. Additionally, the disagreement among researchers regarding the appropriate cutoff score is problematic. Findings have not been reliably consistent. One speculation for the disparity is the varied severity of head trauma. Many studies are inclusive of a variety of etiologies for head trauma varying from epilepsy, brain injury, to healthy participants (Guilmette et al., 1994; Lee et al., 1992; Millis & Kler, 1995). The lack of consensus regarding a
specified cutoff score needs to be resolved to improve the FIT so that it may compete
with stronger measures of malingering.

The 21-Item Test (Iverson, Franzen, & McCracken, 1991) consists of 21 target
nouns in a list that is presented. The respondent is asked to freely recall as many of the
nouns as possible. The respondent is then presented with a distracter list that contains
items used as foils in a recognition task. The recognition task presents the respondent a
forced-choice task that presents two items and the respondent is instructed to indicate
which word was in the target list. Random responding during the recognition task would
result in at least 7 correct responses. Iverson, Franzen, and McCracken (1994) conducted
a discriminant function analysis, using free recall and recognition scores, that yielded an
overall classification rate of 90% of their sample of healthy coached malingerers and
control participants who included psychiatric inpatients, community volunteers, and
individuals seen for a neuropsychological evaluation. Iverson and Franzen (1996) found
that a cutoff score of 9 resulted in 100% sensitivity rate for normal effort control group,
22.5% for memory-impaired subjects, and 69% for experimental malingerers.

Iverson and Franzen (1996) determined that a cutoff of less than 2 on the 21-Item
test yielded a 5% classification rate of malingerers and 100% classification rates for the
normal effort group and memory impaired groups. It also yielded an overall 62%
classification rate with a zero false positive rate. Lastly, they found that <14 on the free
recall section of the logical memory test yielded a 60% classification rate for malingerers,
100% rate for normal effort group, and 95% rate for memory impaired participants, with
an overall 83% correct classification rate and a 2% false positive rate.
Inman and Berry (2002) used a cutoff score of <3 words recalled in free recall with a sample of non-litigating head injured participants who were asked to malinger or respond honestly. Also included in the study was a group of non-head injured normal effort group and a non-head injured group asked to malinger cognitive deficits. The cutoff yielded a 100% specificity rate, but a 0% sensitivity rate with an overall hit rate of 50%. Using a cut of score of <9 words on the recognition score, their data yielded a 100% specificity score, a 5% sensitivity score, and an overall hit rate of 54%.

The FIT and the 21-Item Test are two quick and easy cutoff method assessments derived specifically for the detection of malingering. They both present respondents with stimuli to be recalled a short time after presentation. But they yield low indices of accuracy and even bona fide brain injured respondents can do quite well. Both appear to be best utilized to identify blatant exaggeration and should be administered at the beginning of a test battery before more complex tests are administered. The early placement of these tests in a battery provides less battery-related contextual information to the respondent who can then use that information to malinger more successfully.

**Forced-Choice Method**

The most popular malingering detection method in clinical practice is the forced-choice paradigm. A forced-choice paradigm exposes the subject to a stimulus and subsequently presents the subject with the stimulus as well as a distracter stimulus. Without any prior exposure to the target stimulus, the respondent has a 50% chance of selecting the target stimulus due to the two-alternative forced choice paradigm. Performance significantly below chance is unlikely and not readily seen in brain-damaged subjects, thus below chance performance is highly specific to malingering. The
malingerer who performs significantly below chance would emphasize deficits by purposefully choosing the distracter stimulus, indicating processing and identification of the target stimulus and then choosing the opposite. However, using performance below chance results in extremely low sensitivity to malingering. Given sufficient forced-choice items, scoring significantly below chance is extremely rare. Thus, most forced choice measures have identified cutoff scores that are more sensitive to malingering, but still adequately specific. Forced choice tests are based on binomial probability and therefore many tests yield a “floor effect” as the tests are much easier than they appear. However, patients with bona fide brain damage can score extremely well. Although tests reviewed in the previous section may contain a forced choice component, the following tests are primarily based on the forced-choice method. Forced choice tests are reviewed in Table 1, but are described below in brief.

The Digit Memory Test (DMT: Hiscock & Hiscock: 1989) presents a respondent with 72 items on individual cards. Each item (target stimulus) is a series of 5 digits that the respondent is instructed to study for 5 seconds. Then, after a 5 second delay between presentation and recognition of the stimulus, the respondent is presented with a card containing the target stimulus and a foil and is instructed choose that target stimulus. A face validity manipulation in the test divides the items into sets of increasing delay between the presentation and the recognition task, specifically, 5, 10, and 15 seconds.

The DMT is one of the more researched methods of malingering detection. Researchers have found good indices of accuracy when comparing a variety of different groups including participants motivated to do well and participants motivated to do poorly (Orey, Cragar & Berry, 2000), healthy and neurological patients (Vickery, Berry,
Inman, Harris, & Orey, 2001), and participants with head injury responding honestly and participants with head injury asked to simulate malingering (Inman & Berry, 2002).

The DMT is easy to administer, particularly the computer form, and attempts to mislead the examinee by appearing more difficult than it is. Furthermore, the DMT is one of the more established tests in malingering research. However, its robustness to coaching is unknown.

The Victoria Symptom Validity Test (VSVT: Slick, 1996; Slick, Hopp, Strauss, & Thompson, 1997) is a test specific to malingering and is similar to the DMT. The VSVT is a computerized test that presents a respondent with 48 items over three sets. In each set, 16 target numbers of 5 digits are shown individually. After each item, a blank screen is then presented to the respondent followed by a target number and a foil number. The respondent is to choose the target number and enter the selection on the keyboard. The blank screen, which serves as a retention interval is 5 seconds for the first set of 16 items and is increased to 10 and then 15 seconds for the subsequent groups.

A major difference between the DMT and the VSVT other than number of items is that the VSVT includes both “easy” and “hard” items. Easy items present foils that share no common digits with the target stimulus while in the hard items, digits in the target and foil stimulus are identical but the numbers have the second and third, or third or fourth digits switched. This is to counteract the behavior of using the first and last digits as identifiers of the target stimulus.

The VSVT has been found to have good sensitivity and specificity (Slick, Hopp, Strauss, & Spellacy, 1996). One factor considered with the VSVT by Strauss et al. (2002) is the possible improved prediction ability with the repeated presentation of the VSVT.
The researchers found that intraindividual variability in conjunction with performance information increased the ability of the VSVT to detect malingering as malingerers were less consistent in performance over administrations than those asked to provide their best effort.

In a more recent study, Slick, Tan, Strauss, Mateer, Harnadek, and Sherman (2003) found that the VSVT was insensitive to bona fide memory impairment resulting from neurological illness, providing further evidence that neurological disorder cannot be incontrovertibly identified as the source of low litigant VSVT scores. However, caution should be taken as only six patients were included in this study. Furthermore, robustness of the VSVT to coaching is unknown.

Another widely used test is the Portland Digit Recognition Test (PDRT: Binder, 1993; Binder & Willis, 1991). In this task, the stimuli of 5 digit numbers are presented by the examiner, followed by 5, 15, and 30 second delay, after which two five-digit numbers are presented. During the delay, a distracter task is presented. The subject is instructed to count backwards from a number given by the examiner. The test runs for 72 items, 18 items for 5- and 15- second delays and 36 items in the 30-second delay. Again, this task is to appear more difficult than it actually is in an effort to counteract appearance of face easiness as determined by the examinee.

Contention exists in the research on the PDRT regarding the most appropriate cutoff score and studies reflect results of cutoffs ranging from 39-61 (Binder & Kelly, 1996; Gunstad & Suhr, 2001; Orey et al., 2000; Vickery et al., 2001). These studies yielded sensitivity ranging from 0-100% and specificity ranging from 80-100%. Research has found that test was not robust to coaching (Gunstad & Suhr, 2001).
When the DMT was compared the PDRT, the DMT demonstrated the larger mean effect size of 1.95 compared to 1.26 of the PDRT (Vickery et al., 2001). The DMT also yielded a higher sensitivity rate of 83.4% compared to the PDRT, which was 43.3%. Specificity rates were comparable with DMT as 95.1% and the PDRT as 97.3%. The DMT research yielded a higher average hit rate of 89.4% compared to the PDRT of 71.2%. These results indicate the DMT to be superior to the PDRT.

The Test of Memory Malingering (TOMM: Tombaugh, 1997) presents the respondent with two learning trials of 50-target item line drawings of common objects for a 3 second exposure. Then, the respondent is shown a two-picture panel that contains one target item and a foil and the respondent is instructed to select which picture was in the target series. Feedback is provided after each item. A computerized version has also been developed. Spreen and Strauss (1998) report adequate reliability with internal consistency alpha coefficients of .94, .95, and .94 for trial 1, trial 2, and the retention trial, respectively.

Tombaugh (1997) conducted four validity experiments on the TOMM. The studies yielded high sensitivity and high specificity. The results of the four studies also indicated that the test was not affected by age, education, or bona fide memory impairment of the participants. Rees, Tombaugh, Gansler, and Moczynski (1998) found that the test was unaffected by depression and also yielded high levels of sensitivity and specificity.

Delain, Stafford, and Ben-Porath (2003) found that in the forensic setting, lower TOMM scores of a group of individuals in litigation were associated with reporting more history of head injury, being minimally cooperative or uncooperative during assessment,
and have more diagnoses of antisocial personality disorder. The presence of these additional factors in the context of subnormal TOMM scores in another sample may provide further evidence of malingering.

The Word Memory Test (WMT; Green, Allen, & Astner, 1996) is a computerized test designed specifically to assess poor effort on memory tasks. The primarily effort indices are based on forced choice methodology. The test measures immediate recognition, delayed recognition, delayed free recall, and long delay free recall. Patients with head trauma are able to pass the recognition portions with ease while simulated malingerers and patients suspected of malingering perform less consistently (Green, Lees-Haley, & Allen, 2002; Hartman 2002; Iverson, Green, & Gervais, 1999). Iverson et al. report the WMT is insensitive to brain damage as individuals with severe brain trauma are able to achieve high scores.

Iverson et al. (1999) examined the performance of a variety of samples and found that healthy volunteers and brain-injured patients were able to obtain high scores while the sophisticated malingerers primarily scored below the cutoff. The researchers attribute the two recognition measures of the test (immediate and delayed) and the unexpected 30-minute delayed recognition to greatly contribute the robustness of the WMT to simulated malingering.

Dunn et al. (2003) examined the robustness of the WMT to both coaching and head injury information. There was no interaction found between coaching and information. There were small effects for information found on Immediate Recognition, Delayed Recognition, and Multiple Choice. The coached group performed better than the other groups. Coaching improved performance, yet not to the degree of the respondents
who were asked to perform normal effort. Information regarding head trauma did not
significantly increase performance.

Tan, Slick, Strauss, & Hultsch (2002) compared the utility of three forced choice
tests in the identification of malingering. The TOMM, VSVT, and the WMT were
administered to two groups of undergraduate students. Half of the sample (N=54) was
asked to malinger brain trauma symptoms during administration and their counterparts
were asked perform as if in recovery from a head trauma. The researchers found that the
WMT was able to most differentiate between the groups given traditional cutoffs, while
the VSVT did not catch all of the below chance scorers. However, it is unknown to what
degree the WMT is robust to coaching.

Although there are a variety of tests available for the detection of malingering, the
assessments have limitations. As noted, the assessments are not robust to coaching or the
robustness to coaching is unknown. Given that a large portion of attorneys provide
information to individuals as how to respond in a convincing manner (Wetter & Corrigan,
1995), there is always room for a new contender.

A recent addition to malingering research is the Letter Memory Test (LMT:
Inman, Vickery, Berry, Lamb, Edwards, & Smith: 1998). The LMT presents stimuli
Each target stimulus is presented for subject study for 5 seconds, followed by a 5 second
delay. The subject is then presented with the target stimulus and one or more foils.
Stimulus length varies from 3, 4, or 5 letters and response foils included vary from 1, 2,
or 3 options in addition to the target stimulus. The increase in target stimuli length and
<table>
<thead>
<tr>
<th>Measure - Author(s)</th>
<th>Task</th>
<th>Samples Tested</th>
<th>Findings</th>
<th>Limitation(s) of assessment</th>
</tr>
</thead>
</table>
| **Digit Memory Test - Hiscock & Hiscock (1989)** | 5 digits exposed for 5 seconds target stimulus presented with 1 foil, increasing delay of 5, 10, and 15 seconds, cutoff = <90% | Motivated to do poorly<sup>1</sup>   
Motivated to do well<sup>1</sup>   
Control group<sup>1</sup>   
Honest patients<sup>2</sup>   
Healthy simulators<sup>2</sup>   
Neurological malingering<sup>2</sup>   
Healthy honest<sup>2</sup>   
Healthy simulators<sup>2</sup>   
Head injury history honest<sup>3</sup>   
Head injury simulator<sup>3</sup> | Sensitivity: 50-100%; Specificity: 71-100%; Hit rate: 82-100% | Robust to coaching is unknown. |
| **Victoria Symptom Validity Test - Slick, Hopp, Strauss & Thompson (1997)** | 5 digits presented. Target stimulus presented with 1 foil, increasing delay of 5, 10, and 15 seconds. | Control<sup>4</sup>   
Non-compensation seeking<sup>4</sup>   
Simulator group<sup>4</sup>   
Healthy individuals<sup>5</sup>   
Professionals experienced with head injured patients<sup>5</sup>   
Non-litigating individuals with head injury<sup>5</sup>   
Half simulators, half best effort<sup>5</sup> | Sensitivity: 50%; Specificity: 100% | Performance of 0 -5% considered malingering; 95 – 100% classified as valid. 5 - 95% inconclusive. Robustness to coaching unknown. |
| **Portland Digit Recognition Test - Binder & Willis (1991)** | 5 digits presented. Target stimulus presented with 1 foil after delay of 5, 15, or 30 seconds. During delay, distracter task is presented. | Head injured patients<sup>6</sup>   
Malingers<sup>6</sup>   
Head injury (motivation to perform well, perform poorly, and standard control group)<sup>7</sup>   
Normal effort controls<sup>8</sup>   
Naïve simulators<sup>8</sup>   
Informed simulators<sup>8</sup>   
Warned simulators<sup>8</sup>   
Honest patients<sup>9</sup>   
Healthy simulators<sup>9</sup>   
Neurological malingering<sup>9</sup>   
Honest healthy respondents<sup>9</sup> | Cutoff=39-61: Sensitivity: 0-100%; Specificity: 80-100%; Hit rate 67.2-74.6% | Contention exists regarding most appropriate cutoff score. Test is not robust to coaching. |
<table>
<thead>
<tr>
<th>Measure - Author(s)</th>
<th>Task</th>
<th>Samples Tested</th>
<th>Findings</th>
<th>Limitations of assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test of Memory Malinger ing - Tombaugh (1996)</td>
<td>Presents two learning trials of 50 target item line drawings for three seconds of exposure. Then target stimulus is presented with a foil. Feedback is given after response. Cutoff=90% or below on Trial 2.</td>
<td>Honest healthy individuals&lt;sup&gt;10&lt;/sup&gt;, Neurologically impaired individuals&lt;sup&gt;10&lt;/sup&gt;, Student simulators&lt;sup&gt;11&lt;/sup&gt;, Student honest&lt;sup&gt;11&lt;/sup&gt;, Brain injury simulators&lt;sup&gt;11&lt;/sup&gt;, Brain injury control&lt;sup&gt;11&lt;/sup&gt;, Brain injury litigating&lt;sup&gt;11&lt;/sup&gt;, Brain injury non-litigating&lt;sup&gt;11&lt;/sup&gt;</td>
<td>Sensitivity: 84-100%; Specificity: 95-100%</td>
<td>Research supporting the TOMM includes studies of very limited sample sizes (e.g., N=18, 20, 40)</td>
</tr>
<tr>
<td>Word Memory Test - Green, Allen, &amp; Astner, (1996)</td>
<td>Presents twenty two word pairs. Target word is presented with foil(s). Immediate and delayed recognition are measured. Delayed free recall is also measured.</td>
<td>Mild head injury litigants&lt;sup&gt;12&lt;/sup&gt;, Severe head injury litigants&lt;sup&gt;12&lt;/sup&gt;, Healthy volunteers&lt;sup&gt;13&lt;/sup&gt;, Brain-injured patients&lt;sup&gt;13&lt;/sup&gt;, Sophisticated malingerers&lt;sup&gt;13&lt;/sup&gt;, Coached healthy group&lt;sup&gt;14&lt;/sup&gt;, Informed healthy group&lt;sup&gt;14&lt;/sup&gt;, Coached and informed group&lt;sup&gt;14&lt;/sup&gt;, Control group&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Patients with less severe injuries performed worse than those with severe injuries. When healthy groups were compared, coaching improved performance slightly. Information regarding head trauma did not significantly increase performance. Volunteers and brain-injured patients achieved high scores while sophisticated malingerers primarily scored below the cutoff. Sensitivity 96-100%.</td>
<td>Robustness to coaching not reliably established</td>
</tr>
</tbody>
</table>

Note. <sup>1</sup>Orey, Cragar, & Berry (2000); <sup>2</sup>Vickery, Berry, Inman, Harris, & Orey (2001); <sup>3</sup>Inman and Berry (2002); <sup>4</sup>Slick, Hopp, Strauss, and Spellacy (1996); <sup>5</sup>Strauss, Slick, Levy- Bencheton, Hunter, MacDonald, and Hultsch (2002); <sup>6</sup>Binder & Kelly (1996); <sup>7</sup>Orey, Cragar & Berry (2000); <sup>8</sup>Gunstad and Suhr (2001); <sup>9</sup>Vickery, Berry, Inman, Harris, and Orey (2001); <sup>10</sup>Tombaugh (1997); <sup>11</sup>Rees, Tombaugh, Gansler, and Moczynski (1998); <sup>12</sup>Green, Iverson, and Allen (1999); <sup>13</sup>Iverson, Green, and Gervais (1999); <sup>14</sup>Dunn, Shear, Howe, and Ris (2003)
the increase of foils were meant to counteract subject identification of face easiness or
difficulty which may assist patients in correctly identifying easy items and feigning on
items that appear more difficult. This was designed to address the problem that
individuals may readily realize the task is not as difficult as presented. This is to
compensate for a limitation of the FIT as noted by Iverson and Franzen (1996).

Inman and Berry (2002) administered the LMT to a sample of non-litigating head
injured participants who were asked to malinger or respond honestly. Also included in the
study was a group of non-head injured normal effort group and non-head injured group
asked to malinger cognitive deficits. Using a cutoff score of <93% correct resulted in
100% specificity, 73% sensitivity, and 87% overall hit rate in their sample of head injury
history college student controls and head injury history college student simulated
malingers.

Orey et al. (2000) used the published cutoff score of <93% and yielded a
sensitivity of 58% and a specificity of 100% in their sample of students with a history of
head injury assigned to one of three levels of motivation (motivation to perform well,
motivation to perform poorly, and standard control group). In comparison to the DMT,
PDRT, and LMT, the researchers found the LMT to have the highest sensitivity. Inman
and Berry (2002) found that using the same cutoff score resulted in 100% specificity,
73% sensitivity, and an 87% hit rate in identification of simulated malingerers versus
control, regardless of history of head injury. Malingering groups yielded lower scores
regardless of length and number of response choices. In comparison to the DMT, the
LMT provided a higher sensitivity rate, 73% compared to the DMT sensitivity of 64%.
However, its robustness to coaching has not yet been tested.
Although the LMT is fundamentally a forced-choice method of malingering assessment, it can also be considered a pattern of performance method as, much like the AVLT, some studies have attempted to evaluate patterns of malingered performance on the different levels of difficulty presented to the respondent. Inman et al. (1998) found that the neurological control group, with no known motivation to malinger, demonstrated no significant changes in performance across the blocks of trials that manipulated face-easiness of the test. However, the naïve simulators, coached simulators, and suspected motivationally impaired group demonstrated clear declines in performance as the task appeared to be more difficult, indicating that there may be value in using patterns of performance on the LMT to detect malingering.

Forced choice measures are a prevalent means of assessing if an individual is malingering cognitive deficits. Forced choice tests other than the LMT have some limitations including not being robust to coaching or robustness to coaching is unknown. The results of some studies are based on very small sample sizes that may not be applicable to larger samples. The LMT, however, is a new contender as it is a forced choice test that also gives information on patterns of performance. However, at this point, the LMT has only been examined by its authors. Given this, the LMT warrants further study by an independent research laboratory. Furthermore, it is important to explore additional methods of detecting malingering other than the existing pre-LMT assessments.

The purpose of the present study was to combine the three methods of malingering detection (cutoff score, patterns of performance, and the forced choice method) in examining the utility of the LMT in the identification of respondents asked to
malinger cognitive deficits. The established cutoff score of 93% correct on the forced
choice Letter Memory Test was used in the experiment and patterns of performance were
explored by examining the effectiveness of the two face validity manipulations embedded
in the LMT. To increase the generalizability of the results, simulators were individuals
who reported a prior history of head injury, and the validity of the LMT was tested within
the context of a full neuropsychological battery. To examine the robustness of the LMT
to coaching, some simulators were warned about the possibility of malingering detection
during the evaluation.

Hypotheses of the Present Study

1. The LMT will yield high indices of accuracy (sensitivity, specificity,
positive predictive accuracy, and negative predictive accuracy) when
comparing simulators to persons putting forth normal effort.

2. The LMT will be specific to malingering and not cognitive impairment
secondary to head injury (discriminant validity); there will be no difference
in performance of the head injured normal effort group and the non-head
injured normal effort group on total percent correct on the LMT.
Furthermore, none of the individuals in either of these groups will fall
below an LMT cutoff score provided by prior data (93%).

3. The LMT will be robust to the effects of coaching; there will be no
difference in detection rates of warned versus naïve simulators.

4. The LMT will be a more accurate measure of malingering than the FIT; the
LMT is expected to yield higher indices of accuracy, and the FIT will not
be robust to coaching.
5. The face validity manipulations of the LMT will show differential performance in the two simulator groups when compared to the normal effort controls. It is expected that the simulator groups will perform worse over the levels of difficulty (increasing length of stimulus, increasing number of response choices) than the head injured best effort and healthy best effort groups.
Method

Participants

Participants were selected utilizing the psychology department mass screening process in which introductory psychology students were provided a packet of questionnaires that were voluntarily completed. Included in the packet was a questionnaire for the present study (see Appendix A). Students were asked about any head injury history. Both students with head injury history and students without head injury history were invited to participate in the experiment. Inclusion criteria for the head injury group included a self-reported history of head injury accompanied by a loss of consciousness of at least one minute (or loss of consciousness of any length plus a diagnosis of concussion by a physician). Participants in either group who reported current psychological problems, drug or alcohol abuse, a history of learning disability and/or other significant neurological history were not included in the study.

The total number of participants in the completed study was 102. A random selection of individuals who qualified for the study based on the screen were contacted by phone or electronic mail and invited to participate. Participants who reported a history of head injury were randomly assigned to one of three groups: normal effort, simulation of head injury, and warned simulators. Twenty-five participants were in the normal effort and warned simulators group, each. Twenty-four participants were in the group of participants with a history of head injury who were asked to malingering and given no warning. Finally, one healthy control group consisting of 28 individuals with no history of head injury were included and asked to perform their best. All participants were assigned an identification code, which was used on all materials associated with the
participant to protect confidentiality of records. Upon completion of participation, students were provided with extra credit for the participant’s introductory psychology class (one credit per hour of study completion). No individuals opted to prematurely discontinue participation.

Five additional participants participated in the study but their data was excluded due to non-compliance, which was assessed by their responses on the compliance questionnaire (see Appendix D). The exclusion criteria are discussed further in the section below. Of the five participants excluded, three were in the group asked to malinger but given no warning, one was in the group asked to malinger and was given the warning, and one was in the group of participants who reported a prior head injury with loss of consciousness and asked to perform with best effort.

Procedure

The experimental protocol was administered to participants by the author and trained research assistants under the supervision of a licensed psychologist. Each student was asked to read and sign a consent form (See Appendix B), which was also verbally reviewed with the student by the experimenter. Participants were then provided an identification code to protect confidentiality. As noted previously, participants with a history of head injury were quasi-randomly assigned to one of three groups: normal effort, simulation of head injury, and warned simulators. The testing examiner provided written instructions to the participant. Instructions were in closed envelopes and the envelopes containing the different instructions were randomly intermixed by shuffling them together and having the testing examiner grab an envelope so the examiner was not aware of which set of instructions were being handed to the participant. Near the
completion of the study, only certain experimental conditions were yet to be completed to satisfy chosen sample size. Therefore, not all sets of instructions were being randomly mixed in and thus true randomization did not occur for the full set of participants.

A portion of the head injury history group and the entire non-head injury group were instructed to perform to their best effort. Another portion of the head injury history group was asked to simulate brain damage in a believable way. Finally, the last portion of head injury participants was asked to simulate brain damage in a believable manner, but was also warned that one or more of the tests given was designed to catch faking. Instructions for each group are provided in Appendix C. Testing examiners left the room while instructions were read. The participants were instructed to not discuss the assigned condition with the experimenter.

The participants then took the Letter Memory Test within a full battery of neuropsychological tests (to simulate in vivo conditions of taking cognitive tests in a clinical situation), which were administered according to standard instructions. However, a counterbalanced design was used to administer the 60-minute battery (see Table 2) to determine the significance, if any, of the early or late administration of the Letter Memory Test and the FIT in reference to the respondents’ performance. Half of all participants received the tests in the following order (with approximate administration times): Auditory Verbal Learning Test target list trials, distracter list, and recall of target list (5 minutes), Complex Figure Test copy administration (5 minutes), Letter Memory Test (20 minutes), Auditory Verbal Learning Test 30 minute recall and recognition (5 minutes), Complex Figure Test 30 minute recall (2 minutes), Controlled Oral Word Association (3 minutes), Digit Span (WAIS) (5 minutes), Trailmaking Test, test A and B
(5 minutes), 15-Item test recall (5 minutes), and Auditory Verbal Learning Test 60 minute recall and recognition (5 minutes).

The other half of the participants received the battery in the following order: Auditory Verbal Learning Test target list trials, distracter list, and recall of target list, Complex Figure Test copy administration, 15-Item test recall, Controlled Oral Word Association, Digit Span (WAIS), Trailmaking Test, test A and B, Auditory Verbal Learning Test 30 minute recall and recognition, Complex Figure Test 30 minute recall, Letter Memory Test, and Auditory Verbal Learning Test 60 minute recall and recognition.

Following the test battery, each participant completed a compliance check (see Appendix D) to assure he or she followed the instructions given. Respondents identified by question 1 as not understanding the instructions were excluded. Also if the respondent indicated on question 2 that they did not attempt to follow the instructions at all (a response rating of 1) or only attempted to follow instructions “a little” (a response rating of 2), their data was excluded. At the conclusion of the neuropsychological testing, participants were given credit points for their participation, one point per hour.

Neuropsychological Measures

The Rey Auditory Verbal Learning Test (AVLT: Rey, 1964) measures immediate and delayed verbal learning, span and memory. A 15-item target list is presented to the participant five times, and after each individual trial, the participant is asked to recall the list. Then a distracter list is presented and the participant is asked to recall the list after presentation. Following the distracter list, the participant is asked to recall words from the original target list. After a 30-minute delay, the participant is asked to recall the target
<table>
<thead>
<tr>
<th>Approximate Time (in minutes)</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administration 1: 5</td>
<td>Auditory Verbal Learning Test (AVLT)</td>
</tr>
<tr>
<td>5</td>
<td>Rey-Osterrieth Complex Figure Test (CFT)</td>
</tr>
<tr>
<td>20</td>
<td>Letter Memory Test</td>
</tr>
<tr>
<td>5</td>
<td>AVLT 30 minute recall</td>
</tr>
<tr>
<td>2</td>
<td>CFT 30 minute recall</td>
</tr>
<tr>
<td>3</td>
<td>Controlled Oral Word Association Test</td>
</tr>
<tr>
<td>5</td>
<td>Digit Span (WAIS)</td>
</tr>
<tr>
<td>5</td>
<td>Trailmaking Test</td>
</tr>
<tr>
<td>5</td>
<td>15-item test</td>
</tr>
<tr>
<td>5</td>
<td>AVLT 60 minute recall</td>
</tr>
<tr>
<td>Administration 2: 5</td>
<td>Rey Auditory Verbal Learning Test (AVLT)</td>
</tr>
<tr>
<td>5</td>
<td>Rey-Osterrieth Complex Figure Test (CFT)</td>
</tr>
<tr>
<td>5</td>
<td>15-item test</td>
</tr>
<tr>
<td>3</td>
<td>Controlled Oral Word Association Test</td>
</tr>
<tr>
<td>5</td>
<td>Digit Span (WAIS)</td>
</tr>
<tr>
<td>5</td>
<td>Trailmaking Test</td>
</tr>
<tr>
<td>5</td>
<td>AVLT 30 minute recall</td>
</tr>
<tr>
<td>2</td>
<td>CFT 30 minute recall</td>
</tr>
<tr>
<td>20</td>
<td>Letter Memory Test</td>
</tr>
<tr>
<td>5</td>
<td>AVLT 60 minute recall</td>
</tr>
</tbody>
</table>

Note: Total time for both administrations is 60 minutes
list. In a delayed recognition trial, the examiner presents 50 words (including items from the target and the distracter list) and the participant is asked to identify which words were on the original list.

Spreen and Strauss (1998) reviewed the psychometric properties of the AVLT. Test-retest reliability of the total number of words learned over the five learning trials is high, $r = .77$. Talley (1986) and Vakil and Blachstein (1994) purport Trials 1 and B to be indicative of short-term memory and Trial 5 and delayed recall trials to be indicative of memory storage. It has shown to be sensitive to neurological impairment (Powell, Cripe, & Dodrill, 1991), laterality of brain damage (Malec, Ivnik, & Hinkeldey, 1991) and closed-head injury (Geffen, Butterworth, Forrester, & Geffen, 1994).

The Rey-Osterrieth Complex Figure Test (CFT; Rey, 1941; Osterrieth, 1944; Corwin & Bylsma, 1993) presents the participant with a figure to be reproduced by the participant on a separate sheet of paper. Sixty minutes after the participant has reproduced the figure and the target figure has been removed from the participant’s sight, the participant is asked to reproduce the figure from memory. This test provides information on visuospatial abilities as well as information processing, visuographic copying, visuographic memory, strategy use, and integration difficulties. (Lezak, 1995). The psychometric properties of the CFT are adequate as reviewed by Spreen and Strauss (1998), who cite 6-month test-retest reliabilities range from .76-.89. The CFT has been found to be sensitive to head injury (Leininger, Gramling, Ferrell, and Kreutzer, 1990; Poulton & Moffitt, 1995) and that the degree of hippocampal degeneration in moderate-severe head injured patients was related to percent recall on the CFT (Bigler, Johnson, Anderson, Blatter, Gale, Russo, Ryser, Macnamara, & Abildskov, 1996).
The 15-item Test (Rey Memory Test, FIT, Rey: 1964) presents 15 serial items to the participant who is asked to memorize and recall them. The items are presented in five rows with three characters per line: A, B, C; 1, 2, 3; a, b, c; a circle, a square, a triangle; and Roman numerals I, II, III. Most brain damage patients perform the task easily, as the task purposefully appears to be more difficult which is intended to mislead the malingerer (Lezak, 1995). Total score correct is measured with 15 as the ceiling. Lezak identified 9 as a sufficient cutoff score to differentiate malingerers with honest responders. However, Lee et al. (1992) recommended adjusting the cutoff score to 7 after finding higher cutoffs sacrifice specificity. Interrater reliability is reported by Spreen and Strauss (1998) as adequate, citing 95% agreement on item correct score. This test is easy and quick to administer and is a good measure of blatant exaggeration of cognitive deficits. Total items correct, regardless of placement, generated within the time limit, will measure performance. Validity information was presented above.

The Trail Making Test (Reitan, 1958) present a task in which the participant must connect consecutively numbered circles (1-25) on Part A and connect consecutively numbered and lettered circles on Part B by alternating between the numbers and letters (1-A-2-B…13-L). Both parts are timed and results indicate psychomotor speed and executive function as well as visual conceptual and visuomotor tracking (Lezak, 1995). Spreen and Strauss (1998) judge the psychometric properties of the Trail Making Tests to be adequate, citing test-retest reliability at 6 months to be .98 and .67 for Part A and Part B, respectively. Test-retest reliability at one year is .6 and .8 for Part A and Part B, respectively. Spreen and Strauss (1998) review the validity research for the Trail Making test and found that the test loaded on “rapid visual search,” “visuospatial sequencing”

The *Digit Span* (Wechsler, 1987) is a subtest of the Wechsler Adult Intelligence Scale – III. This task measures short-term memory and attention span (Sattler, 1992). The examiner asks the participant to repeat a series of digits presented to the participant orally. Stimulus length ranges from 2-9 numbers. A second portion of the test asks the participant to repeat the digits in backwards order, an activity of transforming the sequence before restating it. Stimulus length ranges from 2-8 numbers. For both section of Digits Forward and Digits Backwards, each trial consists of two series of digits for each stimulus length sequence. Each item is scored as a one if properly restated. Total score for both Digits Forward and Digits Backwards are used. Internal consistency is reported as .90, averaged across all age, and test retest reliability ranges from .73 to .85, also averaged across all ages (Psych Corp, 1997).

The *Controlled Oral Word Association Test* (COWA; Benton & Hamsher, 1989) assesses word production and fluency. In this assessment, the participants are asked to produce as many words as possible (provided a 60 second time limit) that begin with a given letter. This test is a sensitive indicator of brain dysfunction as it assesses response generation and focused attention abilities. Spreen and Strauss (1998) cite the psychometric properties as highly adequate with a one-year test-retest reliability as $r = .7$
and Lezak (1995) reports good validity with a factor loading to abstract mental operations of .62. Spreen and Strauss (1998) review the validity of the COWA. The COWA has loaded with verbal knowledge (desRosiers and Kavanagh, 1987), naming, problem solving, sequencing, and resisting distraction, (Crockett, 1996). The COWA has been found to have high sensitivity to frontal lobe damage (Miceli, Caltagirone, Gainotti, Masullo, & Silveri, 1981) but also greater sensitivity to left- and bi-frontal impairment (Parks, Loewenstein, Dodrill, Barker, Yoshii, Chang et al., 1988; Perret, 1974; Ruff, Allen, Farrow, Neimann, & Wylie, 1994).

The Letter Memory Test (LMT: Inman, Vickery, Berry, Lamb, Edwards, & Smith: 1998). The LMT is a computer-based task in which a participant is presented with stimuli synthesized from the first 10 consonants of the alphabet (B, C, D, F, G, H, J, K, L, M). Each of the 45 target stimuli is presented for subject study for 5 seconds, followed by a 5 second delay. The subject is then presented with the target stimulus and one or more foils. Stimulus length varies from 3, 4, or 5 letters and response foils included vary from 1, 2, or 3 options in addition to the target stimulus. The increase in target stimuli length and the increase of foils was meant to counteract subject identification of face easiness or difficulty which may assist patients in correctly identifying easy items and feigning on items that appear more difficult. Although the psychometric properties of this test are still being evaluated, Inman and colleagues reported an internal consistency coefficient alpha of .926. In the present study, performance is measured by the nine totals yielded by the varied target stimuli length and varied response stimuli, combined. Also the total percent correct of the 45 trials is measured. Validity information was presented above.
Results

The four groups did not significantly differ on the basis of age \((F(3, 98) = .96, p = .41, \text{ns})\). The groups also did not significantly differ on handedness, \((\chi^2 (3, n = 102) = .53, p = .91, \text{ns})\). However, the groups did significantly differ on the basis of gender, \(\chi^2 (3, n = 102) = 11.23, p = .01\). The group reporting no history of head injury was 32% male, the group who reported a history of head injury and asked to malinger and given a warning (warned simulators) was 28% male, the group who reported a history of head injury and were asked to malinger but given no warning (naïve simulators) was 71% male and the group who reported a history of head injury and were asked to give best effort was 44% male. T-tests were conducted to determine the importance of gender on the outcome variables of total LMT score and total 15 Item Test scores. It was found that the LMT total score was not significantly different based on gender status \((t(100) = -.209, p = .84, \text{ns})\), nor was the 15 Item test total score \((t(100) = -.054, p = .96, \text{ns})\). Demographics are summarized in Table 3.

Performance on the LMT

Total LMT scores by group are presented in Table 4. Analysis of variance indicated that the four groups significantly differed on total LMT score \((F(3, 98) = 34.80, p < .001)\). Follow-up analyses (LSD) indicated that when groups were compared to each other, all comparisons were significant at the .01 level except for the comparison between the two best effort groups.
Table 3

Demographics By Group

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Age Mean (SD)</th>
<th>Handedness</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>No head injury best effort</td>
<td>28</td>
<td>18.93 (.72)</td>
<td>24</td>
<td>9</td>
</tr>
<tr>
<td>Head injury warned simulator</td>
<td>25</td>
<td>20.36 (7.33)</td>
<td>23</td>
<td>7</td>
</tr>
<tr>
<td>Head injury naïve simulator</td>
<td>24</td>
<td>18.79 (.78)</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>Head injury best effort</td>
<td>25</td>
<td>19.08 (.91)</td>
<td>22</td>
<td>11</td>
</tr>
</tbody>
</table>

Note. $SD$ = standard deviation

Table 4

Total LMT Scores by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Total Score Mean (SD)</th>
<th>Min – Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No head injury best effort</td>
<td>28</td>
<td>99.13 (2.66)</td>
<td>88.89 – 100.00</td>
</tr>
<tr>
<td>Head injury warned simulator</td>
<td>25</td>
<td>76.62 (19.70)</td>
<td>40.00 – 100.00</td>
</tr>
<tr>
<td>Head injury naïve simulator</td>
<td>24</td>
<td>64.17 (23.11)</td>
<td>17.78 – 100.00</td>
</tr>
<tr>
<td>Head injury best effort</td>
<td>25</td>
<td>9.82 (.61)</td>
<td>97.78 – 100.00</td>
</tr>
</tbody>
</table>

Note. $SD$ = standard deviation, Min = minimum, Max = maximum
Performance on the 15-Item Test

The average 15-Item Test (FIT) scores and related descriptive data for each group are presented in Table 5. Analysis of variance indicated that the four groups significantly differed on total FIT score ($F(3, 98) = 5.60, p = .001$). Follow-up analyses (LSD) indicated that when groups were individually compared to each other, all comparisons were significant at the .01 level except for two. Similar to the groups’ performance on the LMT, the comparison between the two best effort groups on the FIT was non-significant. Also, the performance between the warned simulators and the naïve simulators on the FIT did not indicate a reliable difference in performance.

Table 5

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Mean (SD)</th>
<th>Min – Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No head injury best effort</td>
<td>28</td>
<td>14.68 (.77)</td>
<td>12-15</td>
</tr>
<tr>
<td>Head injury warned simulator</td>
<td>25</td>
<td>13.56 (2.33)</td>
<td>7-15</td>
</tr>
<tr>
<td>Head injury naïve simulator</td>
<td>24</td>
<td>12.83 (3.06)</td>
<td>4-15</td>
</tr>
<tr>
<td>Head injury best effort</td>
<td>25</td>
<td>14.76 (.83)</td>
<td>12-15</td>
</tr>
</tbody>
</table>

Note. $SD$ = standard deviation, Min = minimum, Max = maximum

Order effects of the LMT and FIT

T-tests were conducted to examine the effect of order by comparing the total LMT and total FIT scores of the two counterbalanced groups for the combined simulator
conditions (naïve and warned) collapsed across all other variables. Results indicate no significant difference on the LMT total score based on whether the participant received an early or late administration of the LMT ($t(47) = .81, p = .42, \text{ns}$) but there was a significant difference on the total FIT score between the group of participants that received the early administration of the FIT and those that received the late administration of the FIT ($t(47) = 31.12, p = .01$). Those participants who received the early administration of the FIT performed more poorly on it than individuals who received the late administration of the test. Total scores on the LMT and FIT based on early or late administrations of the LMT and FIT for the combined simulator groups are presented in Table 6. Further analyses indicated that for the no warning simulator group no order effects were found for the FIT ($t(24) = 1.19, p = .25, \text{ns}$) or the LMT ($t(24) = -.12, p = .91, \text{ns}$). However, the warned simulator group demonstrated order effects on the FIT ($t(22) = 2.76, p = .01$) but not the LMT ($t(22) = 1.16, p = .26, \text{ns}$). As expected, order effects were not found for the two best effort groups on either the FIT ($t(51) = -.81, p = .42, \text{ns}$) or the LMT ($t(51) = 1.73, p = .10, \text{ns}$).

**Hypothesis 1: Sensitivity of the LMT to Malingering**

In examination of the total LMT scores using the specified 93% correct response cutoff to identify the difference between the two simulator groups (warned and naïve) and two best effort groups (head injury and healthy), chi-square analysis was significant, as expected, $\chi^2 (1, n = 102) = 55.48, p < .001$. Given the 93% cutoff, the LMT correctly identified 76% of the simulators (sensitivity) and correctly identified 96% of the best effort (specificity). Also, 95% of the participants who were identified as simulators by the test were true simulators (positive predictive accuracy - PPA) and 81% of those
participants who were identified as best effort, were true non-malingers (negative predictive accuracy - NPA).

Closer examination of the appropriateness of the 93% cutoff in the correct identification of the two best effort groups revealed that out of the 28 participants in the non-head injured best effort group, the cutoff yielded 2 false positives, but there were no false positives in the head injured best effort group.

Table 6

<table>
<thead>
<tr>
<th>Administration</th>
<th>N</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>LMT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early</td>
<td>22</td>
<td>66.97 (23.17)</td>
</tr>
<tr>
<td>Late</td>
<td>27</td>
<td>72.26 (22.42)</td>
</tr>
<tr>
<td>15-Item*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early</td>
<td>22</td>
<td>12.05 (3.21)</td>
</tr>
<tr>
<td>Late</td>
<td>27</td>
<td>14.15 (1.77)</td>
</tr>
</tbody>
</table>

Note. SD = standard deviation, Min = minimum, Max = maximum, *p = .01

Hypothesis 2: Specificity of the LMT to Malingering

To test the discriminant validity of the LMT score in its specificity to simulated malingering and not head injury, the performance of the head injured best effort group and the non-head injured best effort group were compared using a t-test, with means of the total percent correct on the LMT as the dependent variable. As expected, there was no
significant difference found on the performance of the LMT between the two groups, indicating that the LMT is not sensitive to head injury \((t(51) = -1.344, p = .19, \text{ns})\). Furthermore, as noted previously, none of the 25 participants in the head injury best effort group obtained LMT scores below the 93% cutoff score provided by prior data. However, 2 of the 25 participants in the non-head injured best effort group obtained LMT total scores below 93%. Those scores were 88.89% and 91.11%.

**Hypothesis 3: Robustness of LMT to Coaching**

To examine the robustness of the LMT, sensitivity of the 93% cutoff was examined for the naïve malingerers and the warned simulators separately. Chi-square analysis comparing detection rates in naïve simulators and head injured best effort groups was significant, as expected, \(X^2 (1, n = 49) = 35.20, p < .001\). The LMT correctly identified 83% of the naïve simulators and identified all 25 participants in the head injury best effort group as best effort, yielding a specificity rate of 100%. Positive predictive accuracy was 100% and negative predictive accuracy was found to be 86%.

When examining LMT performance in warned simulators and the head injury best effort group, chi-square was significant, as hypothesized, \(X^2 (1, n = 50) = 25.76, p < .001\). The LMT cutoff correctly identified 68% of the warned simulators and 100% of the head injury best effort group. Positive predictive accuracy in this sample was found to be 100% and negative predictive accuracy to be 76%.

Another chi-square analysis comparing detection rates in the warned simulators versus the naïve simulators was non-significant, \(X^2 (1, n = 49) = 1.56, p = .21, \text{ns}\). Thus, the results indicate that about same number of participants in the naïve and warned group
are detected, and about same number are missed. A summary of LMT chi-square values of group comparisons and indices of accuracy using 93% cutoff is presented in Table 7.

Table 7

<table>
<thead>
<tr>
<th>Group Comparison (n)</th>
<th>(X^2) Value</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPA</th>
<th>NPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulators (49) versus best effort (53)</td>
<td>55.48*</td>
<td>75.51</td>
<td>96.23</td>
<td>94.87</td>
<td>80.95</td>
</tr>
<tr>
<td>Head injury naive (24) and head injury best effort (25)</td>
<td>35.20*</td>
<td>83.33</td>
<td>100.00</td>
<td>100.00</td>
<td>86.21</td>
</tr>
<tr>
<td>Head injury warned (25) and head injury best effort (25)</td>
<td>25.76*</td>
<td>68.00</td>
<td>100.00</td>
<td>100.00</td>
<td>75.76</td>
</tr>
</tbody>
</table>

Note. *\(p < .001\), \(X^2\) = chi square, PPA = positive predictive accuracy, NPA = negative predictive accuracy.

Hypothesis 4: Comparison of LMT to the 15-Item Test

Sensitivity, specificity, and robustness to coaching of the FIT was examined and then compared to the LMT. For the FIT, using the cutoff score of 9 correct, chi-square analysis of the head injury best effort and the head injury naïve simulator groups was significant, \(X^2 (1, n = 49) = 4.54, p = .03\). The FIT correctly identified only 17% of the simulators, indicating low sensitivity. However, it correctly identified all 25 of the head injury best effort group, indicating 100% specificity. Positive predictive accuracy was
100%, indicating that all participants identified as simulators were true simulators, but only 56% of those that the test identified as best effort were true non-malingers.

Chi-square analysis comparing detection rates in warned simulators and head injury best effort group was non-significant, \( \chi^2 (1, n = 50) = 3.19, p = .07, \) ns. The cutoff score correctly identified only 12% of the warned simulators, again yielding low sensitivity. The cutoff score correctly identified all 25 of the head injury best effort group, indicating high specificity. Positive predictive accuracy was found to be 100% and negative predictive accuracy to be 53%.

Lastly, chi-square analysis was non-significant when comparing warned simulators with the naïve simulators, \( \chi^2 (1, n = 49) = .218, p = .64, \) ns, indicating no difference in detection rates in the two simulator groups. A summary of FIT chi-square values of group comparisons and indices of accuracy using 9 as a cutoff is presented in Table 8.

These analyses collectively indicate that in the present sample of head injured participants across the groups of naïve malingerers, warned malingerers, and best effort, the cutoff of 9 correct on the FIT was a poor detector of malingering as it yielded low sensitivity and, thus, a high number of false negatives.

The malingering detection rates of the LMT and the FIT, using their respective cutoffs, were then compared using the chi square test for correlated proportions (McNemar, 1975). This analysis compares the classification of participants on both tests but focuses on discrepancies between test predictions. The analysis yielded significant results, \( \chi^2 (1, n = 102) = 12.14, p < .001, \) indicating a significant difference in the detection rates on the two tests. More specifically, the LMT identified all of the
simulators that were identified by the FIT, but the FIT missed 32 of the simulators that were identified by the LMT. This gives further evidence of the relative lack of sensitivity of the FIT in comparison to the LMT.

Table 8

<table>
<thead>
<tr>
<th>Group Comparison (n)</th>
<th>X² Value</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPA</th>
<th>NPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head injury naïve (24) and head injury best effort (25)</td>
<td>4.54*</td>
<td>16.67</td>
<td>100.00</td>
<td>100.00</td>
<td>55.56</td>
</tr>
<tr>
<td>Head injury warned (25) and head injury best effort (25)</td>
<td>3.19</td>
<td>12.00</td>
<td>100.00</td>
<td>100.00</td>
<td>53.19</td>
</tr>
</tbody>
</table>

Note. *p < .05, X²= chi square, PPA = positive predictive accuracy, NPA = negative predictive accuracy.

Hypothesis 5: Patterns of Performance on the LMT

The Letter Memory Test contains two manipulations of face difficulty embedded in the test for the purpose of creating an appearance of an increasing level of task difficulty. One manipulation is the increasing length of the target stimulus (three, four or five letters). The other is the increasing number of choices the subject is presented with from which to identify the target stimulus (two, three, or four choices).
To test the effect of the length manipulation, a repeated measures MANOVA was conducted, with group status as a between subject variable and target stimulus length as a within subjects variable. Descriptive statistics for stimulus levels by group are presented in Table 9.

Table 9
Descriptive Statistics For Stimulus Levels By Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No head injury best effort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 string</td>
<td>98.57</td>
<td>5.55</td>
</tr>
<tr>
<td>4 string</td>
<td>99.52</td>
<td>1.75</td>
</tr>
<tr>
<td>5 string</td>
<td>99.29</td>
<td>2.10</td>
</tr>
<tr>
<td>Head injury warned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 string</td>
<td>78.13</td>
<td>18.21</td>
</tr>
<tr>
<td>4 string</td>
<td>76.80</td>
<td>20.83</td>
</tr>
<tr>
<td>5 string</td>
<td>74.93</td>
<td>24.14</td>
</tr>
<tr>
<td>Head injury naïve*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 string</td>
<td>70.83</td>
<td>23.82</td>
</tr>
<tr>
<td>4 string</td>
<td>63.33</td>
<td>23.01</td>
</tr>
<tr>
<td>5 string</td>
<td>54.17</td>
<td>28.71</td>
</tr>
<tr>
<td>Head injury best effort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 string</td>
<td>99.49</td>
<td>1.85</td>
</tr>
<tr>
<td>4 string</td>
<td>97.33</td>
<td>9.23</td>
</tr>
<tr>
<td>5 string</td>
<td>100.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. *p<.001
The analysis indicated that, as expected, there was an interaction between the length of the target stimulus and group status \((F(6, 102) = 7.486, p < .001)\). To test the nature of this interaction, four repeated measures ANOVAs were conducted, separately for each group. ANOVA results for each group are presented in Table 10.

Table 10

<table>
<thead>
<tr>
<th>Within Subject Effects By Group For Stimulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>------------------------------</td>
</tr>
<tr>
<td>No head injury best effort</td>
</tr>
<tr>
<td>Head injury warned</td>
</tr>
<tr>
<td>Head injury naïve</td>
</tr>
<tr>
<td>Head injury best effort</td>
</tr>
</tbody>
</table>

Note. \(df = \) degrees of freedom

For the head injury best effort group, non-head injury best effort group, and head injury warned simulators, there was no significant within-subject effect, indicating performance across target stimuli length was stable. However, analysis of variance results showed a significant within-subject effect for the head injured naïve simulator group in reference to stimulus length \((F(2, 24) = 15.17, p < .001)\). When the LMT total score means of each stimuli length were compared to each other within this group, all comparisons were found to be significant at the .05 level. As expected, performance worsened as the stimuli length increased for the naïve simulators.
A second repeated measures MANOVA was conducted, with group status as the between subject variable and number of choices as a within subjects variable. Descriptive statistics for the number of response choices by group are presented in Table 11. Results indicated that, as expected, there was an interaction between the number of choices and group status ($F(5, 102) = 4.97, p < .001$).

To test the interaction, four repeated measures ANOVAs were conducted, separately for each group. Similar to findings for target stimuli length, the head injury best effort group, non head injury best effort group, and head injury warned malingering group showed no significant within subject effects. However, for the head injury naïve malingering group, there was a significant within subject effect ($F(2, 24) = 10.75, p < .01$). ANOVA results for each group are presented in Table 12. When the LMT total score means of each response length were compared to each other within this group, all comparisons were found to be significant at the .05 level. As expected, performance worsened as the response length increased for the naïve malingerers.

Figures 2 and 3 illustrate the performances on the LMT by the different groups on the two manipulations of stimulus and choice, respectively.
<table>
<thead>
<tr>
<th>Group</th>
<th>Levels</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No head injury best effort</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 levels</td>
<td>99.05</td>
<td>2.99</td>
<td></td>
</tr>
<tr>
<td>3 levels</td>
<td>99.76</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td>4 levels</td>
<td>98.57</td>
<td>4.57</td>
<td></td>
</tr>
<tr>
<td><strong>Head injury warned</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 levels</td>
<td>80.27</td>
<td>18.61</td>
<td></td>
</tr>
<tr>
<td>3 levels</td>
<td>75.47</td>
<td>21.83</td>
<td></td>
</tr>
<tr>
<td>4 levels</td>
<td>74.13</td>
<td>24.06</td>
<td></td>
</tr>
<tr>
<td><strong>Head injury naïve</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 levels</td>
<td>69.72</td>
<td>21.80</td>
<td></td>
</tr>
<tr>
<td>3 levels</td>
<td>61.39</td>
<td>26.08</td>
<td></td>
</tr>
<tr>
<td>4 levels</td>
<td>57.22</td>
<td>26.95</td>
<td></td>
</tr>
<tr>
<td><strong>Head injury best effort</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 levels</td>
<td>97.33</td>
<td>9.23</td>
<td></td>
</tr>
<tr>
<td>3 levels</td>
<td>99.73</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td>4 levels</td>
<td>99.73</td>
<td>1.33</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $p<.001$
Table 12

Within Subject Effects By Group For Choice

<table>
<thead>
<tr>
<th>Group</th>
<th>Sum of Squares</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>No head injury best effort</td>
<td>20.11</td>
<td>2</td>
<td>10.05</td>
<td>1.966</td>
<td>.150</td>
</tr>
<tr>
<td>Head injury warned</td>
<td>520.30</td>
<td>2</td>
<td>260.15</td>
<td>2.190</td>
<td>.123</td>
</tr>
<tr>
<td>Head injury naïve</td>
<td>1944.44</td>
<td>1.35</td>
<td>1444.93</td>
<td>10.75</td>
<td>.001</td>
</tr>
<tr>
<td>Head injury best effort</td>
<td>96.00</td>
<td>1.063</td>
<td>90.28</td>
<td>1.595</td>
<td>.219</td>
</tr>
</tbody>
</table>

Note. $df$ = degrees of freedom

Figure 2

LMT Percent Correct On Stimulus Length By Group

Note. LMT = Letter Memory Test, Non-HI BE = non-head injury best effort, HI WW = head injury with warning, HI NW = head injury with no warning (naïve), HI BE = head injury best effort.
Non-Malingering Measures

An exploratory analysis of the non-malingering measures included in the assessment battery was conducted. The head injury best effort group did not demonstrate impairments on the non-malingering assessments. No impairments were anticipated due to the group having a history of only mild head injury. The malingering groups, however, did perform worse than the two best effort groups. Therefore, these simulators would have escaped detection had it not been for the inclusion of the LMT in the assessment battery. Table 13 summarizes the performance of the groups on selected non-malingering tests.
Table 13

Group Performances On Non-Malingering Measures

<table>
<thead>
<tr>
<th>Neuropsychological Tests</th>
<th>healthy best effort mean (SD)</th>
<th>best effort mean (SD)</th>
<th>naive mean (SD)</th>
<th>warned mean (SD)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rey Complex Figure Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>copy</td>
<td>31.36 (3.29)</td>
<td>31.16 (2.87)</td>
<td>29.04 (6.16)</td>
<td>30.08 (4.99)</td>
<td>1.45</td>
<td>.23</td>
</tr>
<tr>
<td>recall 1</td>
<td>19.61 (6.07)</td>
<td>20.44 (5.89)</td>
<td>14.98 (5.73)</td>
<td>17.16 (6.73)</td>
<td>4.06</td>
<td>.009</td>
</tr>
<tr>
<td>Auditory Verbal Learning Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trial 5 recall 2</td>
<td>12.23 (1.92)</td>
<td>12.68 (1.75)</td>
<td>8.67 (3.50)</td>
<td>9.48 (2.89)</td>
<td>15.21</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>immediate recall 3</td>
<td>10.50 (2.49)</td>
<td>11.40 (2.52)</td>
<td>5.83 (3.23)</td>
<td>7.48 (2.71)</td>
<td>22.41</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>30 minute recall 3</td>
<td>9.82 (2.76)</td>
<td>10.52 (2.28)</td>
<td>4.83 (3.06)</td>
<td>6.60 (2.63)</td>
<td>24.84</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>30 min recognition 3</td>
<td>14.21 (.99)</td>
<td>14.36 (.91)</td>
<td>10.33 (3.03)</td>
<td>12.20 (2.42)</td>
<td>22.31</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Digit Span 3</td>
<td>12.54 (1.73)</td>
<td>12.80 (2.61)</td>
<td>7.92 (2.45)</td>
<td>10.56 (2.96)</td>
<td>20.93</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Trailmaking Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>22.11 (6.24)</td>
<td>24.24 (7.06)</td>
<td>44.83 (38.60)</td>
<td>33.63 (16.57)</td>
<td>6.15</td>
<td>.001</td>
</tr>
<tr>
<td>B</td>
<td>46.32 (13.00)</td>
<td>50.76 (16.00)</td>
<td>72.33 (36.90)</td>
<td>63.86 (20.62)</td>
<td>6.83</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. Individual comparisons: ¹ Head Injury No Warning significantly worse than both Best Effort. ² Both Simulator significantly worse than both Best Effort. ³ Both Simulator significantly worse than both Best Effort, No Warning significantly worse than Warned. ⁴ Both Simulator significantly worse than No Head Injury Best Effort, No Warning significantly worse than Head Injury Best Effort.
Discussion

The present study examined the utility of the recently developed Letter Memory Test (LMT) in the identification of malingerers of cognitive deficits due to head trauma. Overall, the results of the present study indicate that the LMT yielded adequate indices of accuracy, including sensitivity, specificity, positive predictive accuracy and negative predictive accuracy, in multiple comparisons of groups.

With any measure of malingering of cognitive deficits, it is important to first consider whether if the assessment measure is sensitive to the cognitive changes that might be present in mild head injury. The comparison of the performance of the head injured best effort group and the healthy best effort group indicated that the two groups did not produce reliable differences in performance on the LMT. This suggests that, indeed, the LMT was specific to malingering.

In the present sample, the conventional cutoff score yielded 76% sensitivity, 96% specificity, 95% PPA, and 81% NPA when comparing the two simulator groups to the two best effort groups. Furthermore, the test yielded no false positives in the head injured normal effort group. This is consistent with prior research (Inman & Berry, 2002; Orey et al., 2000), providing further validation for the utility of the LMT in detection of malingering. Unexpectedly, however, two non-head injured best effort participants scored below 93%. The scores were only slightly lower than the cutoff (88.89% and 91.11%). This small discrepancy from the cutoff may reflect poor motivation or test fatigue of the participants. However, both respondents indicated on the compliance questionnaire that they understood the instructions, made an effort to follow the instructions, and believed
they did so reasonably successfully. Therefore, the two low scores may point to issues of specificity, and further research on the LMT will help to clarify this issue.

The LMT versus the 15-Item Test

One purpose of the present study was to compare the utility of the LMT and 15-Item Test (FIT). Prior research indicates that the FIT has low sensitivity but high specificity (Inman & Berry, 2002; Vickery et al., 2001). In comparison of the LMT to the widely used FIT, the present study found the LMT to be highly superior to the FIT in sensitivity and equal in high specificity in reference to the head injury normal effort group and the head injury naïve malingerers. Moreover, the LMT also identified all malingerers identified by the FIT, yet the FIT missed over 30 cases that were identified by the LMT. The difference in sensitivity is likely to be due to the inclusion of two face validity manipulations embedded in the LMT and no face validity manipulations in the FIT.

Also, unlike the FIT, the LMT was superior in that no order effects were found in consideration of the early or late administration of the assessment. Individuals who received the late administration of the FIT have more contextual information regarding the actual difficulty level of the test and are likely able to judge the true difficulty level of the FIT more accurately and respond accordingly. Clearly, prior research indicates that late administration of the FIT in a battery of tests allows the respondent to consider the context of the assessment and feign impairment more convincingly than those respondents with little contextual information. Those with contextual information characteristically pass easy items and fail seemingly difficult items (Iverson & Franzen, 1996). On the other hand, the LMT purposefully presents as more difficult than it actually
is, which misleads the malingering, who thus performs more poorly. These results, in addition to the superior sensitivity of the LMT, indicate that despite the quick and easy administration of the FIT, the LMT proved to be a superior and more useful tool in the identification of malingerers in this sample.

Despite the ease and quick administration of the FIT, it yields low sensitivity to malingering, compromising the utility of the test. In addition, contention exists in research regarding the most beneficial cutoff score for the FIT as the cutoffs of both 9 and 7 have been used (Arnett et al., 1995; Inman & Berry, 2002; Lee et al., 1992; Millis & Kler, 1995). Exploratory analysis using a cutoff of 7 on the FIT with data from the present study yielded extremely low specificity, as only three respondents in the study scored 7 or below on the FIT. One respondent was in the category of head injury with warning and 2 were in the group of head injury no warning. Regardless, cutoffs may vary according to patient population and setting in which the FIT is utilized. Research on the LMT, however, consistently reports 93% correct as the most beneficial cutoff score, regardless of sample.

Effects of Warning

To examine the effects of warning, the performance of the warned simulator group and the naïve simulator group were compared on the LMT, using the 93% correct cutoff. The analysis indicated that LMT was robust to warning, as the detection rates were similar for both the naive and sophisticated groups.

The factor of the two face validity manipulations in the LMT was also explored. The two manipulations include an increase in length of target stimulus and an increase in the number of response choices. Both serve to counteract responder-measured face
difficulty. Results showed that the naïve malingerers performed worse as stimulus length and number of responses increased, while the performance of the warned malingerers and the best effort groups remained relatively stable. Inman and Berry (2002) found similar results, as the simulated malingering groups (head injury and healthy) performed more poorly as the number of choices increased, particularly in the comparison of three and four choices to two choices. Although the researchers also found that the two malingering groups yielded overall lower scores on the LMT than the honest responding groups (head injury and healthy), the two simulator groups yielded consistent performance as stimulus length increased. In other words, stimulus length did not reliably affect scores for the two simulator groups. The slight difference in the results of the two studies is likely due to the difference in experimental group, as Inman and Berry did not include a warned group.

The two malingering groups were comprised of non-litigating participants asked to malinger and healthy participants asked to simulate cognitive deficits.

In future research, it may be fruitful to cross-validate findings with independent samples in order to gather more information of how performance is affected by stimulus and choice length. These patterns of performance can then be used to aid in the clinical identification of malingering. In the present study, for example, the naïve informed malingering group yielded a different pattern of performance than the other groups on the two face-validity manipulations of the LMT, which included increased length of target stimulus and increased number of foils. However, the overall cutoff score yielded high detection rates for both informed and uninformed malingering groups. Malingers tend to perform worse as items appear more difficult. This pattern of performance is worth closely examining in future research to determine how much of a difference in
performance by one respondent over successively more difficult items is diagnostically significant.

Limitations of the Present Study

Part of the inherent difficulty in assessing malingering is identification of the appropriate sample to study (Rogers et al., 1993). Some researchers contend that a simulated malingering design does not adequately mimic the clinical and forensic based motivation of a litigant seeking substantial monetary benefits or other awards and therefore impedes generalizability (Gillis, Rogers & Bagby, 1991), while other researchers report no hindrance to using simulators (Bernard, 1990; Binder & Willis, 1991). In a recent study, Demakis (in press) successfully used the cutoffs on the CVLT and the WAIS derived for malingering detection in patient samples, with college student samples asked to simulate malingering, indicating the use of student simulator samples may be more generalizable to patient samples than once believed. In addition, the present study attempted to at least partially address the limitation of the simulator design with several methods. First, the LMT was administered as part of a battery of tests clinically relevant to the environment in which malingering of cognitive deficits occur. In addition, the present study purposefully used participants who had experience of mild head injury with loss of consciousness and thus should be more familiar with typical head injury symptoms. These individuals are likely to have the ability to malinger in a more believable manner. Regardless, further validation of the LMT as a malingering measure requires its use in a clinical setting using both patients believed to be malingering and psychiatric and medically ill control groups to confirm the usefulness of the LMT in the clinical detection of malingering.
Future Research Directions

Research Paradigms. Debates continue to exist not only over how well simulated malingerers can accurately feign cognitive impairment but whether or not the simulated malingerers have enough motivation to attempt to do so well and in a believable manner. Some researchers have suggested implementing a monetary lottery system to increase the motivation of the participants (Rogers, 1997; Rogers et al., 1993;). A study by Wilhelm, Franzen, Grinvalds, and Dews (as cited in Rogers, Harrell, & Liff, 1993) assigned healthy undergraduate students to one of four conditions: information regarding brain injury, financial incentive ($20) for successful malingering, information and financial incentive, and a control group. All groups were asked to simulate malingering. The researchers found that the group offered financial incentive were more easily detectable on a dot counting test than those individuals who did not receive the incentive.

Bernard (1990) found similar results in a study that compared the memory performance of undergraduate simulators who were either asked to malinger with a financial incentive or without a financial incentive ($50). A control group was also included. The research found that the performance of the two malingering groups did not produce reliable differences, although the performance of both simulator groups was worse than the control group. Although there is little evidence that proves this method useful (Heaton et al., 1978; Inman & Berry 2002; Vickery, 2001) the use of monetary and other incentives continues to be a popular component of simulated malingering research design.

With regard to the LMT, Orey et al. (2000) incorporated a financial incentive into their research paradigm when examining performance of subjects on multiple
motivational and neuropsychological measures, including the LMT. All subjects had a history of head injury. Some subjects were motivated to perform well or motivated to perform poorly via a financial incentive of $25; both groups were compared to a standard head injured control group. However, the researchers found no reliable differences in performance on the neuropsychological or motivational tests between the group with financial incentive to perform well and the standard instruction group. Those individuals who received the financial incentive to perform poorly yielded significant decreases in performance on the tests. Sensitivity was with moderately to moderately high. In regards to the LMT, the researchers found the LMT and two other forced-choice tests to yield high specificity (100%) and moderate sensitivity (58%) in the identification of poorly motivated participants. They found the LMT to be superior to the FIT and 21-Item Test. Furthermore, they found scoring below a cutoff on even one forced choice measure should prompt further investigation. These findings, in conjunction with the findings of the present study, suggest that the LMT is a useful measure of malingering. Future research regarding different kinds and levels of incentives need to be explored to cross-validate findings.

*The Definition of Malingering.* The diagnostic criteria for malingering of cognitive impairment proposed by Slick et al. (1999) may reduce variation of how malingering is diagnosed in both the clinical and the research setting. They stress the importance of external incentive, negative response bias, and the discrepancy of assessments scores with other known data. Furthermore, the diagnostic qualifiers of “definite,” “probable,” and “possible” can provide a more comprehensive concept of malingering. A more unified definition of malingering can help resolve disagreements in
the criteria proposed by the DSM-IV. They report the DSM-IV criteria leaves too much room for error. Slick and co-researchers stress that malingering is under volitional control and has external incentives. This is contrasted with the similar disorders of conversion and factitious disorders, which are associated with psychological incentives. Furthermore, conversion disorder is not under volitional control. Another issue they address is the inappropriate mutual exclusiveness between factitious disorder and malingering and conversion disorder and malingering by the DSM-IV. The researchers purport that comorbidity can occur with malingering and factitious disorder as well as malingering and conversion disorder. Slick et al. report the necessity to be cognizant of these issues and use diagnostic qualifiers to be more accurate in the diagnosis of malingering. Future research on the topic of malingering of cognitive deficits that include the alternate diagnostic criteria by Slick et al. can reduce diagnostic error that is more likely to occur using the DSM-IV diagnostic criteria.

Poor effort can occur with both factitious disorder and conversion disorder. A respondent with either of these diagnoses may score below the cutoff on the LMT and/or the FIT. However, a low score on either or both tests does not warrant a malingering diagnosis in and of itself. Additional information, such as that proposed by Slick et al., is necessary in order to diagnose malingering such as availability of external incentives, pre-trauma functioning, etc. As mentioned previously, malingering is under volitional control; it is purposeful, self-directed behavior. However, this may be more difficult to assess than external incentives or pre-trauma functioning. No specific standard behavioral criterion is available to guide the assessor in making the determination as to whether a
behavior is under volitional control. As with many DSM-IV disorders, making the diagnosis of malingering does include developing clinical impressions and inferences.

*The Daubert Standard.* Research on the detection of malingering is aimed at identifying individuals in the clinical and forensic setting who malinger cognitive deficits and gain clinical and financial resources intended for those individuals with bona fide deficits. Malingerers in the forensic setting can usurp these resources, which may potentially lead to a denial of resources to a deserving individual. Therefore, accurate diagnosis in the forensic setting is of great import. Legal cases have impacted the how malingering assessments and expert testimony is used in the forensic setting.

In 1993, the U.S. Supreme Court set a standard for allowing expert scientific testimony in response to *Daubert v. Merrell Dow Pharmaceuticals* (1993). Vallabhajosula and van Gorp (2001) applied the mathematical standard to tests of malingered cognitive deficits, indicating that a trial judge must consider if the underlying methodology supporting the testimony is scientifically valid. The mathematical standard is complex, incorporates Bayes’s Theorem and takes into account the sensitivity, specificity, prior odds, likelihood ratio, hit rates, and false alarm rate of a given psychological test. Vallabhajosula and van Gorp proposed that a “‘positive’ result on the test must yield a positive predictive value (PPV) of greater than .8, assuming that the pretest probability of malingering equaled .3” (Mossman, 2003). Vallabhajosula and van Gorp found the TOMM to pass this standard but that the FIT test failed to meet the standard.

Mossman (2003) identified 5 criticisms of the Daubert standard. First, the standard only accounts for one of two outcomes and neglects to take into account that cutoff scores can fall under the two overlapping bell shaped distributions of sensitivity
and specificity. Second, the standard does not differentiate between moderately accurate tests and very good ones. Third, due to varying cutoffs in research for malingering tests, some accurate tests can fail the criterion and tests that fail can be made to pass. Fourth, very similar cutoff scores can be treated very differently. For example, Mossman indicates that scores of 44 and 45 on the TOMM have opposite conclusions regarding malingering, positive and negative, respectively. Furthermore, two failing scores can vary greatly (i.e. 44 and 10) yet one indicates much more evidence of malingering. In other words, magnitude is disregarded. Lastly, Mossman points out that the varying base rates of malingering can result in error when applying the Bayesian criterion.

Mossman (2003) indicated that the use of the Daubert standard can be more accurately applied when confidence intervals are taken into account and when expert consensus of base rates is utilized. Currently, there is no research regarding whether or not the LMT meets the Daubert criteria and this is an important area for future research.

Reaction Time. Considering the concept that malingering involves cognitive processes, as the malingerer must consider an appropriate response to a task and then present a less functional response, one branch of research that may be fruitful is to consider including reaction time for responses to memory items. For example, Alban (2003) found that the reaction time measures were sensitive to the added cognitive functions required to complete neuropsychological tests. Malingerers appear to take additional time to identify the correct answer and subsequently chose the wrong answer. This suggests that the added cognitive load makes it more difficult to maintain feigned symptomatology convincingly. Using cutoff scores, Dunn et al. (2003) found that reaction time to items on a computerized task, combined with correct responses, was
helpful in detecting respondents who were not performing to their best effort. Further studies examining the utility of adding reaction time to malingering measures may provide further evidence for the effectiveness of this technique.

Coaching and Warning. Coaching has continued to be a challenge in the area of neuropsychological assessments, as nearly half of lawyers take time to prepare a client for psychological testing and influence performance (Lees-Haley, 1997; Wetter & Corrigan, 1995). Suhr and Gunstad (2000) found reduced sensitivity of a forced-choice memory malingering test in identifying simulators resulting from a mere warning to the respondent that part of the assessment battery is aimed at identification of malingering. This is a fruitful vein of research as it adheres to the ethical principle of protecting the integrity and usefulness of the test in a clinical setting. The present study did indeed protect the integrity of the LMT by only providing a warning to the participant that one or more of the tests was aimed at detecting malingering. The detection rates of the naïve and warned simulators were similar, indicating robustness of the LMT to the warning level of coaching.

Prior research has looked at the effect that different types of information have on performance. Information regarding head injury, information regarding strategies for talking specific tests, combined information on head injury and test strategies, irrelevant information, and no information have all been studied (DiCarlo et al., 2000; Feldstein, Durham, Keller, Klebe, & Davis, 2000; Lamb et al., 1994; Rose et al., 1998; Storm & Graham, 2000; Suhr & Gunstad, 2000;). It appears that information given that includes specific test-taking strategies pertaining to a particular test is more effective for avoiding detection of malingering than less specific information regarding consequences of head
injury. But we have the ethical dilemma of protecting the integrity of malingering and other neuropsychological assessments. A resolution to this is simply incorporating a warning about malingering detection in the assessment battery, as this has been found to influence performance (Gunstad & Suhr, 2001; Suhr & Gunstad, 2000). In sum, future research does not need to violate ethical standards by disclosing information that compromises the integrity of a test. A warning in the battery protocol that indicates to the respondent that one of the tests is designed to identify malingering can satisfy ethical concerns while allowing researchers to further study the effects of warning and coaching. Yet, in the forensic setting, litigants are likely to be coached by legal counsel regarding specific malingering detection avoidance methods.

Multiple Measures of Malingering. Using multiple measures sensitive and specific to malingering can resolve the question of whether poor neuropsychological test performance is due to malingering/poor motivation or reflects actual neurological injury. Also, results of multiple measures can ascertain patterns of inconsistent performance associated with the malingering of cognitive deficits. The comprehensive post-injury information can then be compared to objective pre-injury functioning to determine any bona fide deficits. Contextual factors in which the assessment is given (i.e., litigation, personal injury, financial compensation) needs to be assessed to identify any intent and purpose of malingering deficits. Furthermore, consistent with Slick et al. 1999, qualifying malingeringers into categories of “definite,” “probably,” and “possible” can convey diagnostic certainty by the assessor. The LMT is easily compliant with this multimethod approach to diagnosis.
Accurately identifying malingerers is important in the effort to reduce inefficient external incentives and disability awards. However, over-identifying malingerers can result in the withholding of treatment or awards to individuals who require additional services. Because of these consequences of misidentification, it is prudent to always utilize more than one index in the diagnosis of malingering. The forced-choice method of detection is useful in that yielding a bona fide score significantly below chance is extremely rare, but research continues to examine other factors such as litigation status, co-occurring diagnoses, and cooperativeness (Delain et al., 2003; Suhr & Gunstad, 2002).

In the present study, no participants in the healthy best effort or head injury best effort scored below chance. However, 2 participants in the head injury with warning (8%) scored below chance while 8 participants in the head injury with no warning (33%) scored below chance.

Conclusion

Given the high indices of accuracy that the LMT yielded in the detection of simulated malingerers in the present study and in prior research (Inman & Berry, 2002; Orey et al., 2000), the LMT is shows promise as a competitive assessment for malingering detection in the clinical and forensic setting. The combination of forced-choice and cutoff score features of the LMT yield patterns of performance that can provide further useful information in the detection of malingering.

The psychometric properties of the LMT make it a viable option to include in a neuropsychological assessment battery that may typically be used in the clinical setting. Patterns of performance found within the LMT as well as the assessment battery need to be further explored in various disorders or types of head injury for normative data. This
information will be useful to then apply to the clinical setting. With further cross-validation research with independent samples, patterns of performance can be captured quantitatively to allow for interpretation when used in the clinical setting.

It is important in the forensic setting to accurately detect malingering for the purpose of efficiently doling financial and clinical resources. Unscrupulous attorneys coach litigants how to avoid detection methods. By using multiple measures of malingering, and examining complex patterns of performance, it is more difficult to successfully feign cognitive deficits. The LMT is easily compliant with this multi-method approach to diagnosis.

Further research may shed light on complex response patterns of malingerers and the development of methods to counteract coaching and make it more difficult for sophisticated malingerers to negotiate their way through assessment undetected. The present study, in conjunction with previous research, indicates that the LMT shows promise in the identification of the malingering of cognitive deficits.
References


methods with a large clinical sample. *Psychological Assessment, 6*, 218-224.


Inman, T.H., & Berry, D.T. (2002). Cross validation of indicators of malingering:
a comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology, 17*, 1-23.


detection of malingering: Normative observations on patients with neurological

psychological and neuropsychological cases. *Assessment, 4*, 321-324.

Neuropsychological deficits in symptomatic minor head injury patients after
concussion and mild concussion. *Journal of Neurology, Neurosurgery, and
Psychiatry, 53*, 293-296.

University.


*Psychological Assessment, 3*, 82-88.

Neuropsychological correlates of localized cerebral lesions in non-aphasic brain-


Learning Test in the detection of incomplete effort in neuropsychological


AVLTX: Validation with clinical samples. *Journal of Clinical and Experimental Neuropsychology.*


Appendix A
Mass Screening Questionnaire

Name: ____________________________
Phone: ____________________________
Email: ____________________________
Gender (circle one): Male Female
Current level of education (circle one): Freshman Sophomore Junior Senior
Age: ______________
Which hand do you primarily write with? Left Right

Have you ever experienced a head injury or concussion? Yes No
If YES, when? (date) ______________
Did it involve a loss of consciousness? Yes No
How long did you lose consciousness (in minutes, hours, or days)? __________
What was the length of the posttraumatic amnesia? (i.e. how much time before and after the injury were you awake but you don’t remember what happened?) _______
Were you hospitalized? Yes No
What medications were you given, if any? ______________

Do you currently have a problem with drug or alcohol abuse? Yes No
Do you have a learning disability? Yes No
Do you have any other neurological history other than the head injury? (i.e. seizures, brain tumor) Yes No
Were you involved in any legal litigation regarding your head injury? Yes No
Are you currently receiving any treatment for psychological problems (i.e., anxiety, depression?) Yes No
Please list any current medications: __________________________
Title of Research: Performance on cognitive tests: a simulation study
Principal Investigator: Becca Greub, M.A., Doctoral Candidate
Co-Investigator: Julie Suhr, Ph.D.
Department: Psychology

Federal and university regulations require us to obtain signed consent for participation in research involving human subjects. After reading the statements below, please indicate your consent by signing this form.

The purpose of this study is to contribute to research on how cognitive performance is affected by various instructions. You will be asked to take a number of thinking and memory tests according to instructions provided to you in the envelope. The instructions will provide you with all the directions you need to take the tests. Some of the tests require a written response and some require an oral response, one task is computerized. The typical time needed to complete these tests is two hours. You may discontinue your participation at any time and receive credit for the participation you have completed. One credit is provided per hour of participation.

There are no known direct risks to you by participation in this study and no direct benefits. This research will contribute to our understanding of how certain test instructions affect cognitive performance.

Should you choose to participate, you will be assigned a number that will be used on all your test materials. To protect your confidentiality, your name will not be on any of the testing materials.

Should you have any questions or concerns please speak with the researcher. You may also request to speak with Becca Greub, M.A., bgreub@msn.com or Julie Suhr, Ph.D., at suhr@ohiou.edu, 593-1707. If you have any questions regarding your rights as a research participant, please contact Jo Ellen Sherow, Director of Research Compliance, Ohio University, (740) 593-0664.

I certify that I am 18 years of age or older and that I have read and understand this consent form and agree to participate as a subject in the research described. I agree that known risks to me have been explained to my satisfaction and I understand that no compensation is available from Ohio University and its employees for any injury resulting from my participation in this research. My participation in this research is given voluntarily. I understand that I may discontinue participation at any time without penalty or loss of any benefits to which I may otherwise be entitled.

Signature ____________________________ Print __________________________
Date _________________________________
Appendix C

Instructions to participants

*Best Effort Group:* You are about to take a series of cognitive tests. Please perform to the best of your ability.

*Head Injury Control Group:* You are about to take a series of cognitive tests. Please perform to the best of your ability.

*Head Injury Naïve Malingering Group:* Think about the types of problems that you experienced after your head injury. In this experiment, your job is to make head injury symptoms appear worse on a battery of neuropsychological tests. Imagine that you were in a car accident in which another driver hit your car. You were knocked unconscious, and woke up in the hospital. You were kept overnight for observation. The doctors told you that you experienced a concussion. Try to imagine that a year after the accident, you are involved in a lawsuit against the driver of the other car. If you are found to have experienced significant injuries as a result of the accident, you are likely to receive a bigger settlement. You have decided to exaggerate symptoms of a brain injury in order to increase the settlement you will receive. As part of the lawsuit, you are required to undergo cognitive testing to determine whether or not you have experienced significant brain damage, you are likely to get a better settlement. If the examiner detects that you are faking, you are likely to lose the lawsuit.

You are about to take a series of cognitive tests that would be used in such a situation. I would like you to simulate brain damage, but in a believable way, such that your examiner cannot tell that you are attempting to fake a brain injury.

*Head Injury Warned Malingering Group:* Same as naïve malingering group with the addition of the following: At least one of the tests you will be given is designed to catch you faking, because it is easier than it looks. Be careful.
Appendix D

Compliance Questionnaire

1. What instructions were you given to follow?

2. How hard did you try to follow the instructions you were given?
   1  2  3  4  5
   Not at all  A little  Some  Moderately  Very hard

3. How successful were you in producing the results asked of you in the instructions?
   1  2  3  4  5
   Not at all  A little  Some  Moderately  Very hard