2019

DECREASING THE SUSCEPTIBILITY OF EFFORT MEASURES TO COACHING: THE TOMM AS AN EXEMPLAR

Shannon Kovach
University of Rhode Island, shannonk59@gmail.com

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Kovach, Shannon, "DECREASING THE SUSCEPTIBILITY OF EFFORT MEASURES TO COACHING: THE TOMM AS AN EXEMPLAR" (2019). *Open Access Dissertations*. Paper 899.
https://digitalcommons.uri.edu/oa_diss/899

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

DECREASING THE SUSCEPTIBILITY OF EFFORT MEASURES TO
COACHING: THE TOMM AS AN EXEMPLAR

BY

SHANNON KOVACH

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2019

DOCTOR OF PHILOSOPHY DISSERTATION
OF
SHANNON KOVACH

APPROVED:

Dissertation Committee:

Major Professor David Faust

W. Grant Willis

Leslie Mahler

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2019

ABSTRACT

Malingering has been the focus of numerous studies in forensic psychology and neuropsychology over the last several decades; however, accurate identification of malingers continues to be a primary concern in these fields, particularly in civil cases that come with costly consequences for erroneous identification. Many measures have been developed over the years to identify individuals attempting to mangle on neuropsychological testing; unfortunately, given the vast amount of information available on the Internet that compromises the security of these tests, they may be particularly vulnerable to attempts to coach, or inform, a potential malingeringer on how to avoid being detected by these measures.

A previous study showed that a commonly used measure of effort, the Test of Memory Malingering (TOMM), was susceptible to coaching based on information that was obtained from the Internet (Kovach, 2018). This finding highlighted the importance of identifying ways to either revise existing effort measures or develop new measures to be more resistant to coaching attempts. The current study sought to examine whether increasing the complexity of the TOMM led to variation in level of performance between groups providing differing levels of effort. Participants were assigned to one of four conditions: best effort (control), intermediate effort (fatigued performance), feigning (simulating symptoms of mild traumatic brain injury [mTBI]), and coached feigning (simulating mTBI while trying not to be detected as faking).

The study addressed three main hypotheses: 1) Participants in the feigning group will demonstrate significantly poorer performance on the revised measure than the control group, intermediate effort group, and coached feigning group; 2) Participants

in the coached feigning group will show significantly poorer performance on the revised measure than the control group; and 3) Participants providing intermediate effort will demonstrate significantly poorer performance on the revised measure than the control group.

Results showed that the coached feigning group had the lowest total score on the revised measure ($M = 27.36$), followed by the feigning condition ($M = 31.43$), the intermediate effort condition ($M = 36.86$), and finally, the control group ($M = 41.50$). A one-way between subjects ANOVA was used to determine if total score on the revised TOMM across effort conditions was statistically significant; results of post-hoc tests indicated that participants in the control condition outperformed participants in the two feigning conditions, but they did not perform significantly better than those in the intermediate effort condition.

Results of this study showed that including similar distractor items and varying the number of response options increased the complexity of the TOMM and led to varying levels of performance between groups providing different levels of effort. This suggests that increasing the complexity of a measure of effort may be a useful strategy for decreasing susceptibility of these measures to coaching attempts; however, additional research is needed using various clinical groups with repeated learning and recall trials to confirm the potential utility of this strategy in reducing susceptibility to coaching.

ACKNOWLEDGMENTS

There are several people I wish to acknowledge for their guidance, assistance, and support throughout this process. First and foremost, I would like to thank my major professor, David Faust, for helping me develop the idea for this study and for all his help, advice, and support not only during the dissertation process, but throughout my time as a graduate student. I would also like to express my gratitude to my core committee members, Grant Willis and Leslie Mahler, who have given me incredibly helpful feedback and advice over the past several years as they had graciously served on all of my committees. I also greatly appreciate Diane Coccozza Martins for serving as my defense chair on several of my committees over the years.

I was very fortunate to have a wonderful team of research assistants: Amanda Baron, Hannah Hoyle, and Stefan Schmidt, and I am very appreciative of their help with the data collection process. I am grateful to Lindsey Anderson and Cherie Taylor for allowing me to use the Psychological Consultation Center to conduct my research and for providing assistance when needed. Finally, I would like to give thanks to my family and friends who have given me so much support over the past few years and who I can always turn to when I need motivation or advice.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS.....	v
LIST OF TABLES	vi
CHAPTER 1	1
INTRODUCTION	1
CHAPTER 2	3
REVIEW OF LITERATURE	3
CHAPTER 3	15
METHODOLOGY	15
CHAPTER 4	22
RESULTS	21
CHAPTER 5	25
DISCUSSION	25
APPENDICES	43
BIBLIOGRAHPY	45

LIST OF TABLES

TABLE	PAGE
Table 1. Demographics of the Sample by Effort Condition.....	36
Table 2. Total Scores on the Revised TOMM by Effort Condition.....	37
Table 3. One-Way Analysis of Variance of Revised TOMM Scores by Effort Condition.....	38
Table 4. Tukey HSD Comparisons of Revised TOMM Scores between Effort Conditions	39
Table 5. Sums of Types of Error by Effort Condition	40
Table 6. Error Analysis on A Revised Question with Similar Distractors	41
Table 7. Average Questionnaire Responses by Effort Condition	42

CHAPTER 1

INTRODUCTION

Attempting to exaggerate or feign cognitive impairment in order to obtain financial compensation is unfortunately all too common in legal disputes in the civil arena (Mittenberg, Patton, Canyock, & Condit, 2002). Although this may be unsurprising given the large sums of money that stand to be gained through successful feigning of impairment, it remains an important concern that has attracted the attention of researchers and clinicians hoping to reduce or solve this problem for decades (Wygant & Lareau, 2015). Many measures have been developed to detect exaggeration or feigning during neuropsychological testing; however, individuals may obtain coaching to try to develop a strategy to avoid detection on these measures. *Coached malingering* can broadly refer to any method intended to provide information to individuals seeking to overcome methods to identify malingering. Coaching may be provided by an attorney, or it may be obtained by the individual through their own means, such as an Internet search on common symptoms of the disorder being feigned or how to avoid being detected as faking (Brennan et al., 2009).

A previous study demonstrated that a popular measure of effort, the Test of Memory Malingering (TOMM), may be susceptible to coaching based on information obtained from the Internet (Kovach, 2018). This finding highlights the need to create new measures, or revise existing effort measures, that are more resistant to coaching attempts, especially given the amount of information on the Internet threatening test

security (Bauer, & McCaffrey, 2006). One such method to reduce susceptibility to coaching might be through increasing test complexity as many popular measures of effort rely on a simple detection strategy than may be easily circumvented through coaching or instruction. More complex measures that rely more heavily on genuine cognitive abilities, such as attention and memory, may create differing levels of performance among healthy and impaired individuals, thus requiring a malingerer to adopt a more complex strategy to avoid detection.

CHAPTER 2

REVIEW OF LITERATURE

Malingering is an important problem in neuropsychological evaluation that has been the focus of numerous research studies over the past few decades. Malingering has been defined as the intentional exaggeration or fabrication of symptoms of injury or illness in order to obtain an external reward (American Psychiatric Association, 2013). In civil cases, this reward is typically monetary compensation, which can run into the millions of dollars. In criminal cases, defendants may be seeking a “not guilty by reason of insanity” ruling, and in these and other cases may fake symptoms of mental disorder to support their claims, argue for mitigating factors, or obtain lesser sentences. Thus, the presence of secondary gain, e.g., seeking some type of external reward, is often evaluated when undertaking a neuropsychological evaluation (Heilbronner et al., 2009).

Malingering continues to place a significant burden on the legal system, even though it has been a highly researched and debated problem in forensic psychology and neuropsychology for many years. As important as it is to detect individuals who are attempting to feign or exaggerate a disorder, it is perhaps more important to avoid falsely identifying someone as malingering when they are not, or to verify true disorder. Falsely labeling an individual as a malingerer can result in the loss of needed compensation in civil cases or in unjust sentencing outcomes in criminal cases (Berthelson, Mulchan, Odland, Miller, & Mittenberg, 2013; Heilbronner et al., 2009).

Thus, it is crucial that neuropsychologists, as well as others who may be involved in malingering status determinations, make every effort to reach accurate conclusions and have the highest quality methods available to assist them.

Mittenberg et al.'s (2002) study estimated that 29% of personal injury, 30% of disability, 19% of criminal, and 8% of medical cases involved probable malingering. A review of malingering in medicolegal contexts estimated the base rate of malingering in chronic pain patients with secondary gain to be 20 to 50% (Greve, Ord, Bianchini, & Curtis, 2009). Although these are merely estimates, they emphasize the prevalence of exaggeration and feigning in the legal system. Contributing to the serious consequences of malingering is the resulting financial burden. Chafetz and Underhill (2011) examined Social Security disability claims for 2011 and estimated the cost of malingering for that year to be 20 billion dollars. This signifies a massive financial burden, especially given that this estimate only focused on a select portion of malingering cases (i.e., Social Security disability claims) for a single year.

Embedded Effort Measures. Determinations of malingering in psychology and neuropsychology are often made through a combination of subjective impressions and the use of assessment measures. Numerous measures to detect faking or exaggeration on neuropsychological testing have been developed to assist in the identification of malingering. Some of these measures are embedded within standard cognitive assessment instruments, for example the forced-choice recognition trial on the California Verbal Learning Test, Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) or the validity scales on the Minnesota Multiphasic Personality Inventory, Second Edition (MMPI-II; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer,

1989). Embedded measures are often intended as a check on an examinee's effort or reporting style during an evaluation. In cases in which effort or cooperation is an especially important concern, such as evaluations for legal cases, it is often advised that a neuropsychologist incorporate more than one indicator of effort, including not only embedded measures, but at least one standalone test (Heilbronner et al., 2009).

Standalone Effort Measures. Standalone tests of malingering are designed primarily or solely to appraise effort or symptom validity. Many standalone measures of malingering have been developed and studied using civil, criminal, psychiatric, and cognitively impaired samples. These tests vary in their psychometric properties, with some measures, such as the TOMM (Tombaugh, 1996) and the Word Memory Test (WMT; Green, Allen, & Astner, 1996) found to demonstrate high levels of sensitivity to feigning or exaggeration, often greater than 80%, while maintaining high levels of specificity (Green, Lees-Haley, & Allen, 2002; Schroeder et al., 2012). Conversely, other effort tests, such as the Rey 15-Item Test (Rey, 1941), have been shown to be highly transparent, and thus to have low sensitivity (Strutt, Scott, Shrestha, & York, 2011).

Ceiling Effects. Standalone measures of effort are often prone to *ceiling effects*, which is a term used to describe the clustering of too many individuals at a test's upper limit (e.g., making few or no errors). Ceiling effects limit the measurement of variation in performance. A relatively high percentage of individuals obtain maximum or near maximum scores on effort tests such as the TOMM and the WMT, which stems from the main detection strategy these tests employ. The tests are designed to appear more difficult than is truly the case in order to detect suboptimal effort, whereas they are

actually very easy. Consequently, most individuals providing sufficient effort, even those with various psychiatric and cognitive impairments, usually make very few or no errors. In contrast, individuals who make considerably more errors than is plausible given their purported injury or illness are often identified as providing insufficient effort (Green et al., 2002; Schroeder et al., 2012).

Effort tests with low ceilings generally cannot differentiate between good effort and adequate effort. For example, the TOMM can detect very poor effort, but may be insensitive to subtle changes or decline in effort (Schroeder et al., 2012). The ability to assess degrees of effort during an evaluation could be valuable to neuropsychologists practicing in a various settings as it could help determine whether an individual was exerting maximal effort, or whether they were simply putting forth just enough effort to get through testing, but not performing at or near the best of their abilities.

This is an especially important determination in cases of ambiguous effort or seemingly low motivation. If an individual does not appear to be engaged in testing or doing their best, or if they have conflicting or borderline performance on indicators of effort, it would be beneficial to evaluate the degree of effort, and not be limited to the type of artificial, dichotomous categorizations associated with many current malingering tests. If an individual is providing modest effort on testing, but not their best effort, then the results obtained during the evaluation may not reflect their true level of ability or potential maximal performance (An, Zakzanis, & Joordens, 2012). Such distinctions can be especially important whether the results are used to assist in legal determinations, or in clinical settings to help design interventions (e.g., medication for ADHD, accommodations at work or school, etc.), perform differential

diagnosis, or predict outcomes. For example, in a legal case, even a subtle decline in cognitive functioning may have a substantial impact on the performance of demanding occupational tasks (e.g., flying commercial aircraft). Consequently, differentiating between a drop in test scores that reflects true brain dysfunction as opposed to less than full effort can change outcomes completely.

Coached Malingering. The potential for an individual to be coached on testing is another factor complicating the assessment of true effort. *Coached malingering* is a term used to describe an individual who obtains information on testing or on the symptomology of the injury or illness they are attempting to feign in order to avoid detection. These individuals may gather information on their own or be provided with information by their attorney or others. A conscientious attorney might warn clients about measures neuropsychologists may use to evaluate feigning or exaggeration, and which could result in false-positive identifications and potential denial of just compensation (Brennan et al., 2009). Worries about expert bias or selection of faulty measures might feed such concerns and action steps. Similarly, truly injured individuals may seek out coaching given concerns they will not be compensated fairly by a legal system they perceive as potentially biased or flawed. Hence, such individuals may be injured *and* exaggerating their level of dysfunction. Nevertheless, whether an injury is totally fabricated or is exaggerated to a degree, it certainly can be argued that it is beneficial to identify whether, or the extent to which, claims are falsified.

Most studies on coached malingering have found many effort measures to be largely resistant to the effects of coaching; however, these studies are prone to

methodological errors or shortcomings. In one such study, the participants who were instructed to malingering were provided with coaching instructions for the Validity Indicator Profile (VIP), but then administered the TOMM, which relies on a completely different strategy to detect malingering (Powell, Gfeller, Hendricks, & Sharland, 2004). A recent study created a set of coaching instructions designed to reflect the amount of information on a malingering measure readily available on the Internet (Kovach, 2018). In this study, using specific coaching instructions developed based on a brief Internet search on the TOMM, every participant in the specific coaching group was able to identify and pass the TOMM on testing. In addition, all participants in this group performed poorly on at least one of the genuine cognitive tests, thus reflecting their ability to follow the instructions to feign dysfunction while avoiding detection on a measure of effort. The results of this study contradicted previous literature on coached malingering by showing that information that could be obtained during a brief Internet search was sufficient to successfully malingering while avoiding detection on the TOMM.

Although the aforementioned study only examined the effects of coaching using information from the Internet on one test of effort, it is likely that other effort tests, especially those using a similar strategy as the TOMM, may be compromised by obtaining information available on the Internet or through other sources. Several effort tests including the TOMM, the Dot Counting Test (DCT), and the WMT are designed to look harder than they actually are, which allows for easy detection of naïve malingerers who have no familiarity with the test or its underlying detection strategy. However, an individual may well avoid detection if they are familiar with the test

design: they only need to provide adequate effort to avoid being caught. With the rise of accessibility to the Internet, a potential malingerer may only require a matter of minutes to obtain sufficient information (coaching) to pass these measures.

Researchers have shown that the security of several effort measures, including the TOMM, the WMT, and the Victoria Symptom Validity Test has been compromised by various websites (Bauer & McCaffrey, 2006; Ruiz, Drake, Glass, Marcotte, & van Gorp, 2002). It is all but impossible to limit the accessibility to information about effort tests on the Internet; thus, further avenues to make tests more resilient to coaching need to be explored.

Complexity of Effort Tests. Increasing the complexity of effort tests is one possible way to increase a measure's resistance to coaching. As previously mentioned, several measures of effort are prone to ceiling effects, which make them poor or limited indicators of degrees of effort. Effort tests that are designed to be more complex may result in producing differing levels of performance depending on an individual's level of effort. For example, an individual, healthy or not, who is fatigued, bored, or otherwise not engaged in testing will generally perform at a lower level than individuals maintaining good effort throughout the process. Whether intentional or not, it is important to identify suboptimal levels of effort because variations in effort during a neuropsychological battery can affect performance on genuine cognitive tests and lead to erroneous conclusions (An et al., 2012).

Test designs aimed at increasing the complexity of effort tests have improved test sensitivity to varying degrees. For example, attempts to reduce the transparency of the Rey 15-Item Test by increasing the perception of task difficulty have been largely

unsuccessful in increasing sensitivity to levels comparable with other standalone measures of effort (Strutt et al., 2011). To illustrate, Boone, Salazar, Lu, Warner-Chacon, and Razani (2002) attempted to improve the sensitivity of the Rey 15-Item Test by adding a recognition trial after the free recall trial. The additional trial did increase sensitivity significantly, but not to the level other commonly used measures of effort typically achieve.

The WMT is an effort measure designed to be more complex than previous effort tests (Green et al., 1996). This test assesses an individual's ability to learn 20-word pairs (e.g., dog-cat) presented either orally or on a computer. Immediate and delayed paired recall trials present options of word pairs (e.g., cat-rabbit), are intended to measure effort, and often seem to demonstrate high levels of sensitivity (Batt, Shores, & Chekaluk, 2008; Green et al., 2002).

The WMT includes four different tasks that vary in difficulty, thereby reducing the measure's transparency, and expanding appraisal beyond just effort alone to also capture aspects of verbal memory capacities. Few studies have assessed the utility of the WMT as a measure of verbal memory apart from its use as an effort test, though one study demonstrated differences in level of performance across the four memory subtests between cognitively healthy controls and brain injured patients (Green et al., 2002). Although Green and colleagues were successful in designing a measure of effort and memory with subtests that vary in difficulty, some studies have shown poor specificity rates for the WMT (Batt et al., 2008; Greve, Ord, Curtis, Bianchini, & Brennan, 2008).

Number of Response Options. There is often agreement across studies assessing the optimal number of response options on multiple choice tests that having three response options is ideal. Reducing the number of options from four or five to three reduces item difficulty and increases item discrimination (Haladyna, Downing, & Rodriguez, 2002; Rodriguez, 2005; Vyas & Supe, 2008). However, further reducing the number of response options from four or five down to two significantly reduced item difficulty as well as reliability of the item and was not advised among many of the authors of these studies; as previously stated, three response options was recommended across most studies (Rodriguez, 2005).

Although varying the number of response options will not impact the difficulty of the item for individuals who have the background knowledge to correctly answer the question, adding more response options decreases the probability that someone who has incomplete knowledge or does not know the right answer, all else being equal, will select the correct answer. With a two-choice response option, the odds of correctly guessing the answer are 50%, but these odds drop to 25% when faced with a four-choice response option. Without varying the difficulty of the question, someone who is randomly guessing will get more questions right with two-choice response options than with three- or four-choice response options (Baek & Wojcieszak, 2009).

Random Order of Item Difficulty. Randomizing the order of item difficulty may make it more challenging for a potential malingerer to monitor their performance during the test. Many tests with items at varying levels of difficulty begin with easy items and become progressively harder. Cognitive or achievement tests are often designed this way so that an individual's confidence will improve after successfully

answering easy items before they get to more challenging items (Skinner, 1999). This design is also efficient as discontinuation criteria can be set so that once an individual fails several items in a row, they are not administered more difficult items that they would likely miss, to decrease testing time and potential frustration. Although having items increase in complexity progressively may be an effective test design for individuals providing good effort, individuals who are not engaged in testing may give up when faced with challenging items and respond randomly. Similarly, individuals attempting to feign impairment may intentionally fail harder items to give the appearance of cognitive deficits (Frederick, 2002).

Several studies have found that varying item difficulty on cognitive testing in a non-progressive manner does not impact the performance of healthy individuals significantly (Skinner, 1999), although most of these studies utilize healthy, cognitively normal samples. Few studies have attempted to examine the performance of individuals who are attempting to provide insufficient effort; based on a review of the literature, these studies are exclusively focused on the VIP (Frederick & Crosby, 2000, Frederick, Crosby, & Wynkoop, 2000).

The VIP (Frederick & Crosby, 2000) is a measure of effort that was designed to have a hierarchy of item difficulty that is randomized throughout the test. Items for the nonverbal section of the measure were modified from the Test of Nonverbal Intelligence (TONI; Brown, Sherbenou, & Johnsen, 1982), whereas the test developer created the items on the verbal section. Given random ordering, it is cognitively challenging for a potential malingerer to select a level at which to start failing items that comports systematically with level of item difficulty (Frederick et al., 2000;

Frederick, 2002). The test generates a score profile that is analyzed for markers of valid versus invalid performance patterns; invalid patterns include careless or irrelevant responding and malingering. A malingering performance curve would show lower than chance (50%) accuracy for the easy items, and then rise to 50% as items increase in difficulty and malingerers can no longer intentionally get questions wrong, but rather have to guess (Frederick & Crosby, 2000). Although the VIP has proven to be useful in identifying insufficient effort, it is limited in its ability to accurately classify intentional underperforming from other types of invalid profiles; it has varying levels of acceptable to poor sensitivity for identifying malingering (73.5% for nonverbal section and 67.3% for the verbal section; Frederick, 2002).

The premise of increasing the complexity of an effort test, especially in a manner that is difficult to track for the test taker, is potentially useful for reducing susceptibility to coaching. In addition, if overly simple malingering measures such as the Rey 15-Item Test are highly transparent as a test to detect effort, more complex tests of effort may have lower transparency given the inclusion of genuinely challenging items. Randomly distributing item difficulty would require potential malingerers to develop a sophisticated strategy in which they keep track of their performance on items of varying difficulty. Failing easy items at or below chance level indicates some degree of random or careless responding or intentional underperforming. Therefore, a potential malingerer would need to get enough easy items right to avoid raising suspicion. At the same time, getting too many items right may categorize their performance as cognitively normal (or above expectations for the type or level of impairment at issue), and thus undermine attempts to feign

dysfunction. Utilizing a hierarchy of item difficulty randomized throughout the test may hold promise for making current measures more resistant to coaching.

The proposed study included four participant groups with varying levels of effort: controls (best effort), intermediate effort, feigning, and coached feigning. All participant groups were administered a measure of effort that was revised to reduce susceptibility to coaching attempts. As this was a preliminary examination of revisions to this measure, initial analyses were primarily aimed at testing the impact of these changes on group performance. This study sought to test the following hypotheses: 1) Participants in the feigning group will have significantly poorer performance on the revised measure than the control group, intermediate effort group, and coached feigning group; 2) Participants in the coached feigning group will have significantly poorer performance on the revised measure than the control group; and 3) Participants providing intermediate effort will have significantly poorer performance on the revised measure than the control group.

CHAPTER 3

METHODOLOGY

Participants. The study was advertised to University of Rhode Island undergraduate students in psychology and related fields, primarily those enrolled in introductory psychology courses. Students in many of these courses are encouraged to participate in research for extra credit. Fifty-six participants from undergraduate classes volunteered to participate in the study. One participant misunderstood the demands of the study and showed up to participate two weeks in a row; as soon as this mistake was detected by the research assistant running the participant, the participant's second set of data was removed from the final dataset to ensure independence of observations.

The final sample size included in the analyses was 56. Most participants ($n = 49$) were between the ages of 18 and 21, with 6 participants between the ages of 22 and 26; 1 participant did not report their age. The majority of participants identified as white ($n = 40$), with 5 identifying as black, 7 as Hispanic, 2 as Asian, 1 as multiracial, and 1 as other. Thirty-three participants identified as female, 19 identified as male, and 2 identified as non-binary; 2 participants did not report their gender.

Procedures. The current study sought to assess whether increasing the complexity of a popular measure of malingering led to variation in level of performance between groups providing optimal effort and differing levels of suboptimal effort. This was an experimental study with one independent variable,

level of effort, which had four levels: best effort (control), intermediate effort (e.g., fatigued performance), feigning (simulating symptoms of mild traumatic brain injury [mTBI]), and coached feigning (simulating mTBI while trying not to be detected as faking). The group of controls was instructed to provide their best effort on the test. The group providing intermediate effort was asked to take the test as if they were very tired after a long day of work or classes. The feigning group received instructions to fake symptoms of mTBI while taking the test; they were provided with common symptoms of mTBI, e.g., headaches and difficulty concentrating. The coached feigning group received instructions that were similar to the feigning group, but in addition they were coached on the purpose of the measure and instructed to perform less than their best, but not to an extent that was so obvious they would be detected as faking (Appendix A). Participants were randomly assigned to conditions prior to the start of testing.

Participants were tested individually in a quiet room on campus by one of three undergraduate research assistants. All three research assistants were blind to the study hypotheses, that is, they were not informed about which groups were predicted to perform better or worse on the revised measure. Upon arriving for the study, participants were given a consent form explaining their rights as a research participant (e.g., their right to refuse to participate or leave at any point). Consent forms were stored separately from participants' test data and questionnaires to ensure anonymity. Participants were administered a revised version of the TOMM followed by a brief questionnaire (Appendix B).

The G-Power 3.1 computer program (Faul, 2009) was used to conduct an a priori power analysis. When reported in the literature, malingering studies typically yield large effect sizes due to extreme differences in scores between malingering and genuinely responding participants. Thus, most malingering studies can be sufficiently powered using relatively small samples. As it was not known how groups would perform on the revised measure prior to conducting the study, the effect size had to be estimated; the estimated effect size was set at a relatively conservative level of 0.50, which is significantly smaller than effect sizes obtained in many previous malingering studies, including Kovach's (2018) previous study, which yielded a very large effect size of ($f = 1.32$). Using a power level of .80, the estimated total sample size needed to detect an effect size of this magnitude was 48 participants. A post-hoc power analysis revealed a large effect size ($f = 0.90$) and an observed power level of 0.99, indicating that this study was sufficiently powered.

Measure. The *Test of Memory Malingering* (TOMM; Tombaugh, 1996) is a 50-item visual recognition test that was developed to distinguish poor effort from adequate effort. The TOMM has two learning trials, administered one right after the other, and an optional retention trial presented after a brief delay. Examinees are shown 50 pictures of common objects, presented one after another. After the set is completed, examinees are shown a series of 50 two-choice recognition items: one answer choice is a previously shown item and the other answer choice is an item that was not (Tombaugh, 1996). Validation studies on the TOMM have found high levels of sensitivity and specificity in distinguishing genuine responders from malingerers using the cut-off score for performance on Trial 2 that is recommended in the test

manual (Haber & Fichtenberg, 2006; Wisdom, Brown, Chen, & Collins, 2012).

Studies have also demonstrated that the TOMM is relatively insensitive to the effects of age, education, and various types of cognitive and psychiatric disorders (Haber & Fichtenberg, 2006; Rees, Tombaugh, Gansler, & Moczynski, 1998).

A recent study showed that the TOMM was susceptible to coaching using information obtained from a brief Internet search (Kovach, 2018). The current study sought to revise the instrument by increasing its complexity and making it more difficult for potential malingerers to track their performance throughout the measure. One revision was to incorporate more response options; instead of relying on two-choice answer options, three- and four-choice answer options were introduced. A second revision was to vary details of the target item to create a similar but different distractor item. In the standard version of the TOMM, the distractor items are vastly different from the target items. For example, if the target item is a picture of a teddy bear, the distractor item is a completely unrelated object, such as a flower.

The previously described strategy of slightly altering target and distractor items is utilized in the Shape Learning Trials of the Neuropsychological Assessment Battery (NAB; Stern & White, 2003), a measure of visual learning and memory. Research suggests that the Shape Learning Trials of the NAB are sensitive to severity of traumatic brain injury and that brain injured individuals perform significantly worse than healthy individuals (Donders & Levitt, 2012). Thus, utilizing a similar design of slightly varying distractor items from target items on the TOMM, at least for select items, was predicted to increase item difficulty and thereby provide one potential means for assessing degree of effort as opposed to limiting judgments or

classifications into binary categories. It is important to note that increasing the difficulty of items will place more demand on genuine cognitive abilities, such as attention and visual memory, than the original version of the TOMM. This is similar to the WMT, which serves primarily as a measure of effort, but includes additional, more difficult, trials that assess verbal memory ability.

Randomizing the order of items introduced a third dimension that potential malingerers had to track, which was intended to make the TOMM more resistant to coaching attempts. Given the limits of human cognition and working memory capacity (7 units plus or minus 2 for unidimensional variables according to Miller's classic (1954) article), adding these additional dimensions to a task would likely create extreme or excessive mental load. A person's performance on a given task is a reflection of the mental load of the task and the mental effort, or cognitive capacity, the person is able to exert (Kirschner, 2002). Thus, on the revised measure, a malingerer would have to keep track of three varying dimensions of complexity while attempting to perform in a manner consistent with individuals who have the genuine illness or injury they are purporting to have, a task which seemingly exceeds the limitations of human cognition for a very large percentage of the population.

Varying the order and level of complexity (e.g. two-, three-, and four-choice answer options) of items not only adds to the mental load of the measure, but also allows a neuropsychologist to evaluate the pattern of performance for inconsistencies or suspicious responding. For example, if an individual is missing several easy items (e.g., items in which the distractors are very dissimilar from the target), but correctly

answering more challenging items at a similar or greater rate, it could signal intentional underperforming.

The proposed measure maintained the two sample items from the original version of the TOMM, followed by the presentation of the same 50 stimulus items for three seconds each, in accordance with standard TOMM administration procedures. As also accords with standard TOMM procedures, 50 forced-choice recognition questions were presented next, although with modifications made in the content of most of these items. Six of these 50 questions were maintained in their original form (two-choice answer options that included a markedly dissimilar distracter option). Among the items that remained, 14 items contained markedly dissimilar distractor options in accord with the original TOMM format, but seven of these items were modified from two-choice to three-choice answer options, and the other seven items from two-choice to four-choice answer options. The 30 additional items contained one distractor option that was similar to the target item (e.g., a shaded flower versus an unshaded flower); ten of these items were two-choice answers, ten were three-choice answers, and ten were four-choice items. Items were randomly ordered throughout the recognition section with the intention of making it more difficult for potential malingerers to keep track of their performance throughout the test.

This was an initial study designed to examine the revised measure's potential effectiveness in detecting coached malingerers. Only one trial of the test was administered to perform a critical check on whether manipulating difficulty across multiple dimensions would change performance characteristics, with an initial focus on accuracy across different conditions. The long-term, programmatic interest is to use

the current research to examine the impact of modifications to the measure, to make revisions and adjustments to this or other measures as needed, and to then conduct subsequent studies to test effectiveness in evaluating effort among various groups (i.e., individuals with TBI, depression, and other neurologic and psychiatric disorders).

CHAPTER 4

RESULTS

Statistical Analyses. Preliminary analyses were conducted using SPSS version 25 (IBM Corp., 2017). Group comparisons indicated that demographic variables (i.e., age, race/ethnicity, sex) were similarly distributed across effort conditions (see Table 1). Statistical assumptions of analysis of variance (ANOVA) were examined, including independence of cases, normality, and homogeneity of variance. No assumptions were violated.

Table 2 provides the mean total score for the revised TOMM for all four conditions. The control group had the highest score ($M = 41.50$, $SD = 3.35$), followed by the intermediate effort condition ($M = 36.86$, $SD = 5.53$), and then the feigning condition ($M = 31.43$, $SD = 7.82$). The coached feigning group had the lowest mean score out of all conditions ($M = 27.36$, $SD = 7.18$), indicating that they had the poorest performance and made the most errors on the revised measure out of all the groups.

Hypotheses 1, 2, and 3 were initially examined using a one-way between subjects ANOVA to determine if total score on the revised TOMM differed significantly across effort conditions. Table 3 presents these results, which reveal that scores differed significantly between effort conditions ($F(3,52) = 13.88$, $p < .001$). In order to determine which groups performed significantly different on the revised measure from one another; Tukey post-hoc tests were examined. The control condition was found to be significantly different at the $p < .05$ level from the feigning and coached feigning

conditions, and the intermediate effort condition was found to be significantly different from the coached feigning condition; all other conditions were not significantly different from one another (see Table 4).

Results of post-hoc tests indicate that participants in the control condition outperformed participants in the two feigning conditions, but they did not perform significantly better than those in the intermediate effort group. Although participants in the intermediate effort condition outperformed those in the coached feigning group, which had the lowest performance, they did not have significantly different performance from either the control group or feigning condition. This indicates that there was less clear delineation between groups demonstrating best effort, intermediate effort, and impaired effort as compared to the original version of the TOMM.

Eta-squared was calculated to determine effect size for the ANOVA. Cohen's (1988) benchmarks were used to interpret eta-squared; results indicated that there was a large effect size for effort condition on the revised TOMM ($\eta^2 = 0.45$). This finding indicates that 45% of the variance in revised TOMM performance was attributable to effort condition. The results suggest that the level of effort participants put into their performance on the revised TOMM accounted for a large amount of the variance in total score on the revised measure.

Additional exploratory analyses were conducted to determine whether the number and type of errors made on items varied by effort condition (see Table 5). Results of these analyses showed that participants tended to make far more errors on questions with similar distractor items compared to items with dissimilar distractors. The type of response option with the lowest mean number of errors was the two-choice, dissimilar

distractor items ($M = 5.7$), which were the six items taken directly from the original version of the TOMM and not altered in any way. When participants were given three- and four-choice response options instead of only two, they were more likely to make errors. The highest mean number of errors were made on three-choice ($M = 22.7$) and four-choice response option items ($M = 25.7$) with similar distractors.

Although asking participants to intentionally feign impaired performance could partially account for the large number of errors made on the revised measure, even participants in the control group missed several items. Of note, controls tended to answer items with similar distractors incorrectly, yet missed very few items with dissimilar distractors, indicating that the items with similar distractors were more difficult than items containing only dissimilar distractors. Table 6 shows the error analysis of an item that was revised to be a four-choice response option with a similar distractor and two dissimilar distractors. Results of this analysis indicate that none of the controls selected the dissimilar distractors; however, some participants in the other conditions did select the dissimilar distractors as their answer.

Participants' responses on the follow-up questionnaire indicated that most rated the instructions they received as very clear, with mean group ratings ranging from 1.14 (feigning group) to 1.86 (coached feigning group) on a 7-point scale with lower scores indicating greater clarity and higher scores indicating little to no clarity. In addition, participants had approximately equal ratings of test difficulty and success at following instructions across conditions; however, the control condition rated their perceived success more highly than the other conditions ($M = 6.29$), which is likely due to the simplicity of their instructions (see Table 7).

CHAPTER 5

DISCUSSION

Determination of malingering can be an onerous task that carries a significant financial and personal burden if a misclassification occurs in either direction, i.e., someone who is genuinely injured being labeled as a malingerer and losing out on much needed financial assistance, or conversely, someone who is fabricating injury winning a large damage award and obligating a party to payment that may far exceed what is warranted. Although there may be little sympathy felt towards large corporate entities paying damage awards, these payments are often passed on to consumers, and beyond that, the person blamed for causing the (pseudo) injury may feel deep personal remorse about the seeming harm they inflicted on another person.

Accurate classification of malingerers has received considerable attention in the field of neuropsychology over the years (Heilbronner et al., 2009; Van Oorsouw & Merckelbach, 2010; Young, Jacobson, Einzig, Gray, & Gudjonsson, 2016), with much progress having been made in the field. However, problems and critical concerns remain. One such central and potentially vexing problem is identifying malingerers who have been coached on how to pass measures designed to detect faking, especially given the ever-increasing ease with which individuals can obtain information about measures of effort and symptom validity using the Internet (Bauer & McCaffrey, 2006; Ruiz et al., 2002).

Kovach's (2018) study showed the susceptibility of a measure of effort to coaching attempts using information available on the Internet. In that study, participants in one condition were given a warning about a measure designed to detect feigning and a brief description of the measure's detection strategy. This minimal coaching allowed all participants in that condition to avoid detection on this effort test. In addition, they were able to distinguish the measure from genuine memory tests and selectively underperform on the latter, thereby creating a result suggesting good effort and genuine memory deficit. The results of that study demonstrated the ease with which information available on the Internet regarding neuropsychological evaluations and commonly used measures can be used to successfully develop an effective coaching strategy, a result unlikely to be isolated to a single malingering test. The current dissertation served as a follow-up to a previous study by attempting to make a popular measure of effort more resistant to coaching attempts.

Results of the current study partially supported the first hypothesis; participants in the feigning group ($M = 31.43$) had significantly poorer performance on the revised measure than the control group ($M = 41.50$). However, they did not have significantly different performance from the intermediate effort ($M = 36.86$) or coached feigning groups ($M = 27.36$). This indicated that participants in the uncoached feigning condition performed well below the level of controls, but not significantly worse than participants exerting intermediate effort and those who were coached on the measure. This suggests that the revised measure cannot as easily distinguish between adequate effort, naïve feigning, and coached feigning. The original version of the TOMM effectively distinguished between adequate effort and intentional feigning using a

cutoff score for total correct on Trial 2 and the retention trial. However, only one recall trial was used in this study, so it is not possible to assess whether a similarly set cutoff point would maintain its effectiveness on the revised measure. Although the results suggest it may be more difficult to differentiate between effort levels using a single cutoff score, this would need to be tested in future studies using the full three recall trials of the measure.

The second hypothesis of the study was supported; participants who were coached on how to feign and given a warning to avoid being detected by the measure performed significantly worse than the control group. In contrast to the results of Kovach's (2018) study the coached feigners had significantly poorer performance compared to controls as well as an intermediate effort group. This suggests that either the coaching instructions provided to the participants were not sufficient to help them develop a strategy to avoid detection or that the revised measure was more resistant to coaching attempts given the increased complexity. In the previous study, coached participants were told that making more than a few errors was highly unusual even among genuinely injured individuals. However, coached participants in this study did not receive such explicit directions regarding number of errors as the average number of errors among healthy controls was not yet known.

The coached feigning group was instructed to do less than their best but not perform so poorly as to make it obvious they were underperforming, their mean total score was the lowest out of all four groups, though only significantly worse than the coached and intermediate groups. Their very poor performance may have been due to a lack of explicit coaching instructions that prevented this group from developing an

effective coaching strategy to “pass” the measure and the instructions may have been too vague to sufficiently coach participants to avoid detection. Future studies should test a new coached feigning group that can be explicitly coached using information regarding the average performance of controls in this study to determine whether the measure remains resistant to more specific coaching attempts or not.

In addition, future studies could identify a sample of “skilled malingerers”, participants who are able to avoid detection on various measures of effort, including the original version of the TOMM. These participants could then be explicitly coached on this revised measure to see if they are can avoid detection, even given detailed information about the measure and the detection strategies being employed. As previously mentioned, the three revisions to this measure create three dimensions of complexity, which seemingly exceeds the limit of human cognitive capacity. Thus, it would be expected even with the most explicit coaching possible, a potential malingerer would not be able to simultaneously track all three dimensions to successfully develop a strategy that mirrors the performance of those who have genuine injury or illness.

Finally, the third hypothesis of the study was not supported; participants who were instructed to provide an intermediate level of effort did not have significantly poorer performance on this measure compared to controls. Although the intermediate group had a lower mean total score than controls, this difference did not reach statistical significance. This suggests that participants providing less than their best effort made more errors compared to controls, but not so many as to perform significantly worse.

One reason the third hypothesis might not have been supported in the current study is due to uncertainty among participants on how to follow the intermediate effort instructions. Although most participants rated the instructions they were given as clear, participants might have experienced confusion over how to translate feeling fatigued into test performance. Some participants may have already been tired at the time of testing and decided to perform the best they could given their already fatigued state, whereas other participants may have intentionally tried to underperform and make more errors than they would have if they were doing their best. Subsequent studies should consider including items on the questionnaire to assess an individual's level of fatigue prior to testing as well as open-ended questions to gather information about how they approached the task to control for differences in performance strategy. Future studies should attempt to recruit a real-life intermediate effort condition to determine how individuals not instructed to perform less than their best, but who for various reasons may not be able to provide maximal effort at the time of testing, perform on this revised measure.

The results of this study indicated that there may be some ability to determine degrees of effort using the revised measure because the binary classification of acceptable versus poor effort no longer holds up with these groups. It is likely that the intermediate effort group would have passed the original TOMM while exerting the same level of effort due to the minimal cognitive demands of the task; however, on this revised measure, they are not easily distinguishable from either the controls or the feigning group, indicating that the more complex measure results in less clear delineation between groups.

The types of errors made on a single item was briefly explored in Table 6; however, further analysis on the average number and type of errors made (e.g., number of errors made on items with similar versus dissimilar distractors) between various groups should be examined in subsequent studies as a potential method of effort classification. Using number of errors made on items with dissimilar distractors may be a more effective classification method between groups rather than a cutoff score for total number of errors made. Utilizing error analysis as a method of classification may also help to ensure the measure is truly classifying degree of effort rather than degree of impairment; that is, analyzing the types of errors made may prevent false positive identifications of poor effort being made among individuals who are attempting to provide their best effort but who cannot perform as well as healthy controls due to genuine injury or illness.

On the original version of the TOMM, individuals with genuine illness and injury perform similarly to controls, allowing the measure to easily screen out anyone performing so poorly it is likely only attributable to low effort. On the revised version, there seems to be greater variation in performance between best effort, diminished effort, and feigned impairment as evidenced by the wide range of scores obtained between groups and the overlap of these ranges between groups (i.e., lowest score obtained by a control was 36, the highest scores in the two feigning groups were 40 and 41). A brief analysis of the type of errors made on a single item indicated there may be a way to further analyze differences in performance between groups as a way to classify effort rather than a single cutoff score classifying effort as either good or poor. As mentioned in the introduction, the ability to determine degrees of effort can

be very useful to neuropsychologists practicing in a variety of settings, not only in the forensic realm, as it allows them to determine if an individual is putting forth maximal effort, if they are providing adequate effort but not doing their best, or if they are displaying poor effort or potentially malingering.

An exploratory analysis across conditions revealed that participants had much higher rates of error on items with similar distractors. This was especially important in the control condition, as healthy controls on the original version of the TOMM made very few, if any, errors. On the revised measure, controls made very few errors on items with dissimilar distractors, but missed several items containing similar distractors. This suggests that changing the similarity of distractor items to more closely resemble target items was successful in increasing the difficulty of those items.

Varying the number of response options did not have as noticeable of an effect on the difficulty of the measure, as participants had approximately equal rates of errors among two-, three-, and four-choice response options; however, these differences were not tested for statistical significance and adding more response options might have a significant effect on the measure's complexity. The items with the least amount of errors were the dissimilar, two-choice response options, which were the items taken directly from the original version of the TOMM. This is not surprising given the low number of errors made by healthy controls and various clinical groups on the original TOMM. The finding of increased number of errors made on other types of items (i.e., ones with similar distractors) indicate that the revisions made to the TOMM successfully increased the difficulty of the measure.

A follow-up study will examine the individual items on the revised measure in greater detail, for example, by formally examining the difficulty of each item. In addition, internal consistency will be analyzed to determine if participants within each group are performing similarly across items. Some items may be further revised or altered to change their difficulty level based on the results of these analyses. For example, an item that most participants are answering incorrectly might be revised to make it slightly less difficult. Such analyses were not included in the current study given the primary intent of determining whether the revisions made to the TOMM led to varying levels of performance across groups. In-depth item analysis and fine-tuning of the measure will be the next step in this program of research.

There are several strengths and limitations to this study. One strength was including a participant group to simulate an intermediate level of effort; most effort measures only discriminate between good/adequate levels of effort and very poor or below chance levels of effort, and one goal for this study was to be able to assess degrees of effort beyond the binary good versus poor effort. Participants attempted to provide less than maximal effort without intentionally trying to feign impaired performance, which allowed for examination of a potential difference in overall performance compared to groups providing good or poor effort. Although the intermediate group did not perform significantly worse than the control group, they did, on average, have poorer performance and made more errors. Another strength of the study was having research assistants who were blind to the hypotheses of the study, so as not to unintentionally bias their interaction with participants.

One limitation in the current study was only testing one recall trial of the revised measure, which prevented the researcher from assessing whether performance would improve in subsequent trials after participants were aware of the attention and visual memory demands of the study. Participants in the control group made significantly more errors on this revised measure compared to the performance of healthy controls on the original version of the TOMM; however, because only one trial was administered, the researcher was not able to determine if controls (as well as other participant groups) would adopt a better strategy for remembering target items and make fewer errors on subsequent trials.

The current study permits one to draw only limited inferences about external validity due to the demographics of the participant sample being predominantly young, white, and female. Although this may be representative of psychology courses at the university, it is likely not reflective of the population of individuals who malingers in real life. This study is also limited in external validity as it did not include a sample of participants with genuine TBI to see how they would perform on this measure. Future studies should replicate and expand this study with more diverse samples who more accurately represent the demographics of populations involved in litigation and who may well have an incentive to malingers.

Moving forward, it is critical that these initial or subsequent findings be cross-validated with various multicultural populations. A variety of multicultural groups, in particular racially and ethnically minoritized groups, might well be at increased risk for overidentification of malingering given a number of factors, such as the long history of findings indicating overpathologizing among such individuals (Faust,

Ahern, Bridges, & Yonce, 2012). Alternatively, in some cases, false-negative errors might be increased, leading members of marginalized groups to be mislabeled as cognitively normal versus genuinely injured, and consequently missing out on much needed compensation or other benefits. In any case, knowing more about generalization to diverse groups would be a welcome development and obviously superior to speculation. In this particular case, it is imperative that the types of revisions to the TOMM being developed and examined here be tested with various multicultural groups to determine efficacy and the potential need for modifications in measurement or interpretation.

Another limitation of the study was the use of an original questionnaire that was developed to collect information about the revised measure. This questionnaire was developed due to the lack of existing measures that would adequately assess the questions the researcher was most interested in addressing; however, the use of an original, non-validated questionnaire could lead to errors in assessing constructs of interest due to poorly worded questions or a range of other psychometric shortcomings.

The results of this study show that including similar distractor items and varying the number of response options increased the difficulty of the TOMM and led to varying levels of performance between groups providing different levels of effort. A coached feigning group that was informed about the purpose of the measure and instructed to perform in a way so as not to get caught faking performed significantly worse than groups providing good or adequate effort. This indicated that increasing the complexity of a measure of effort such as the TOMM may be a useful strategy for

decreasing the susceptibility of these measures to coaching attempts. However, considerable additional research is needed to assess the functionality of the measure with more explicitly coached feigning groups before this can be determined.

Further research should test this revised measure in various groups of interest, for example, groups of individuals with genuine TBI, potential or suspected malingerers, groups with psychiatric disorders that may affect level of effort, etc. Testing this measure in groups with various neurologic and psychiatric conditions may permit the development of cut points that differentiate clinical groups providing good or maximal effort, but who perform worse than controls, from those exerting lower levels of effort. In addition, future studies should test the measure with a second learning and recall trial as well as a retention trial, in accordance with standard administration procedures for the original version of the TOMM, to determine if total scores improve after a second presentation of the stimuli. It is possible that after a second learning trial, the total score for the control group, and possibly other groups, might improve and even pass the recommended cutoff score for the original version of the TOMM. However, further testing is needed to determine whether certain groups improve their total score after a second learning trial or whether they continue to make the same types or number of errors.

TABLES

Table 1

Demographics of the Sample by Effort Condition

	Control	Intermediate Effort	Feigning	Coached Feigning
Age				
Mean	19.77	19.93	20.36	19.64
SD	1.50	1.33	2.73	0.84
Race/Ethnicity				
White	10	12	11	7
Black	1	1	1	2
Hispanic/Latino	1	0	1	5
Asian	1	1	0	0
Multiracial	0	0	1	0
Other	1	0	0	0
Gender				
Female	8	9	8	8
Male	5	4	4	6
Non-Binary	1	0	1	0
Not Reported	0	1	1	0

Note. The follow-up questionnaire included other demographic options that were not endorsed by any participant (*Native American* was included as an option for race/ethnicity, while *Other* was included as an option for gender). SD = Standard Deviation

Table 2

Total Scores on the Revised TOMM by Effort Condition

	Mean	SD	Minimum	Maximum
Control	41.50	3.35	36	46
Intermediate Effort	36.86	5.53	26	45
Feigning	31.43	7.82	16	40
Coached Feigning	27.36	7.18	17	41

Note. SD = Standard Deviation.

Table 3

One-Way Analysis of Variance of Revised TOMM Scores by Effort Condition

	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
TOMM					
Between groups	1607.57	3	535.86	13.88	.000
Within groups	2007.86	52	38.61		
Total	3615.43	55			

Note. SS= sum of squares; df= degrees of freedom; MS= mean square.

Table 4

Tukey HSD Comparisons of Revised TOMM Scores between Effort Conditions

Group 1	Group 2	Mean Diff.	Std. Error	95% Confidence Interval	
				Lower Limit	Upper Limit
Control	Intermediate Effort	4.64	2.35	-1.59	10.88
	Feigning	10.07*	2.35	3.84	16.30
	Coached Feigning	14.14*	2.35	7.91	20.38
Intermediate Effort	Control	-4.64	2.35	-10.88	1.59
	Feigning	5.43	2.35	-0.80	11.66
	Coached Feigning	9.50*	2.35	3.27	15.73
Feigning	Control	-10.07*	2.35	-16.30	-3.84
	Intermediate Effort	-5.43	2.35	-11.66	0.80
	Coached Feigning	4.07	2.35	-2.16	10.30
Coached Feigning	Control	-14.14*	2.35	-20.38	-7.91
	Intermediate Effort	-9.50*	2.35	-15.73	-3.27
	Feigning	-4.07	2.35	-10.30	2.16

Note. * $p < 0.01$

Table 5

Sums of Types of Error by Effort Condition

	Similar Distractor			Dissimilar Distractor		
	2-Choice (n=10)	3-Choice (n=10)	4-Choice (n=10)	2-Choice (n=6)	3-Choice (n=7)	4-Choice (n=7)
Control	38	37	37	0	6	1
Intermediate Effort	39	44	58	7	13	13
Feigning	50	72	80	9	24	25
Coached Feigning	64	74	82	18	39	40
Total	191	227	257	34	82	79
Avg. Number of Errors	19.1	22.7	25.7	5.7	11.7	11.3

Note. Average Number of Errors was calculated by dividing the total number of errors made per type of response option by the number of items of that type.

Table 6

Error Analysis on A Revised Question with Similar Distractors

Answer	Response Type	Control	Intermediate Effort	Feigning	Coached Feigning
Correct	Target	11	10	2	6
Incorrect	Similar Distractor	3	2	10	5
	Dissimilar Distractor	0	1	2	3

Note. Values in cells show how many participants in each group answered the question correctly or incorrectly; if they answered incorrectly, it shows what type of error was made.

Table 7

Average Questionnaire Responses by Effort Condition

	Control	Intermediate Effort	Feigning	Coached Feigning
Clarity of Instructions	1.64 (1.30)	1.36 (1.08)	1.14 (0.36)	1.86 (1.46)
Difficulty of Test	3.43 (1.16)	3.21 (1.42)	3.57 (1.22)	4.07 (1.0)
Perception of Success	6.29 (0.82)	5.0 (1.18)	4.64 (0.84)	5.29 (1.38)

Note. Values reported are group means with the standard deviations in parentheses. Responses were rated on 7-point scales, with clarity of instructions, lower scores indicate greater clarity; with difficulty of test higher scores suggest greater perception of difficulty; with perception of success higher scores reflect greater perceived success at following task instructions.

APPENDICES

Appendix A. Participant Instructions.

Control Group: Please do your best on this test. It is important that you pay close attention to the test items and do the best that you can.

Intermediate Effort Group: Take this test as if you are very tired after a long day of work or classes and it is a little harder than usual for you to focus.

Feigned Mild TBI Group: Take this test as if you recently suffered a blow to the head and sustained a mild head injury or concussion. People with concussions often have serious problems with headaches, and with thinking, concentrating, and especially remembering things.

Coached Feigning Group: Take this test as if you recently suffered a blow to the head and sustained a mild head injury or concussion. People with concussions often have serious problems with headaches, and with thinking, concentrating, and especially remembering things. However, the measure you are about to take is designed to catch people who are trying to fake impairment. You will want to do less than your best, so that you do not appear healthy, but avoid doing so poorly that it is obvious you are faking.

Appendix B. End of Study Questionnaire.

Demographics

Age

Race/Ethnicity

- White
- Black
- Hispanic/Latino
- Asian
- Native American
- Multiracial
- Other

Gender

- Female
- Male
- Non-binary
- Other

Questions

How difficult was this test?

1	2	3	4	5	6	7
Very Easy						Very Difficult

How clear were the instructions you were given at the start of this study?

1	2	3	4	5	6	7
Very Clear						Not at all Clear

What did the instructions you were given at the start of the study ask you to do?

- Provide my best effort on this test
- Take this test as if I was very tired after a long day
- Fake symptoms of mild traumatic brain injury on this test
- Fake symptoms of mild traumatic brain injury and avoid being detected as faking

How successful do you think you were at following the instructions you were given?

1	2	3	4	5	6	7
Not Successful						Very Successful

BIBLIOGRAPHY

- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5. Washington, D.C: American Psychiatric Association.
- An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology*, 27(8), 849-857.
- Baek, Y. M., & Wojcieszak, M. E. (2009). Don't expect too much! Learning from late-night comedy and knowledge item difficulty. *Communication Research*, 36(6), 783-809.
- Batt, K., Shores, E. A., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society*, 14(6), 1074-1080.
- Bauer, L., & McCaffrey, R. J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the Internet: Is test security threatened? *Archives of Clinical Neuropsychology*, 21(1), 121-126.
- Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury*, 27(7-8), 909-916.
- Brennan, A. M., Meyer, S., David, E., Pella, R., Hill, B. D., & Gouvier, W. D. (2009).

- The vulnerability to coaching across measures of effort. *The Clinical Neuropsychologist*, 23(2), 314-328.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1982). TONI, Test of Non-verbal Intelligence: A Language-Free Measure of Cognitive Ability. Pro-ed.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.
- Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology*, 28(7), 633-639.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd edition. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test—2nd edition: Manual*. San Antonio, TX: The Psychological Corporation.
- Donders, J., & Levitt, T. (2012). Criterion validity of the Neuropsychological Assessment Battery after traumatic brain injury. *Archives of Clinical Neuropsychology*, 27(4), 440-445.
- Faust, D., Ahern, D. C., Bridges, A. J., & Yonce, L. J. (2012). Assessment of malingering and falsification: Conceptual foundations and sources of error. In Reynolds, C., & Horton Jr, A. M. (Eds.), *Detection of malingering during head injury litigation* (pp. 1-67). New York, NY: Springer Science & Business Media.
- Frederick, R. I. (2003). Review of the Validity Indicator Profile. *Journal of Forensic*

- Neuropsychology*, 2(3-4), 125-145.
- Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior*, 24(1), 59-82.
- Frederick, R. I., Crosby, R. D., & Wynkoop, T. F. (2000). Performance curve classification of invalid responding on the Validity Indicator Profile. *Archives of Clinical Neuropsychology*, 15(4), 281-300.
- Green, P., Allen, L., & Astner, K. (1996). *Manual for Computerised Word Memory Test*. Durham, NC: CogniSyst.
- Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Archives of Physical Medicine and Rehabilitation*, 90(7), 1117-1126.
- Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist*, 22(5), 896-918.
- Haber, A. H., & Fichtenberg, N. L. (2006). Replication of the Test of Memory Malingering (TOMM) in a traumatic brain injury and head trauma sample. *The Clinical Neuropsychologist*, 20(3), 524-532.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-333.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., &

- Conference Participants 1. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23(7), 1093-1129.
- IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.
- Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, 12, 1-10.
- Kovach, S. (2018). *The effect of coaching on the ability to identify and pass a measure of insufficient effort* (Unpublished master's thesis). University of Rhode Island, Kingston, RI.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1094-1102.
- Powell, M. R., Gfeller, J. D., Hendricks, B. L., & Sharland, M. (2004). Detecting symptom- and test-coached simulators with the test of memory malingering. *Archives of Clinical Neuropsychology*, 19, 693-702.
- Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, 10(1), 10-20.
- Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique.

Archives of Psychology, 28, 286–340.

- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3-13.
- Ruiz, M. A., Drake, E. B., Glass, A., Marcotte, D., & van Gorp, W. G. (2002). Trying to beat the system: Misuse of the Internet to assist in avoiding the detection of psychological symptom dissimulation. *Professional Psychology: Research and Practice*, 33(3), 294-299.
- Schroeder, R. W., Buddin Jr, W. H., Hargrave, D. D., VonDran, E. J., Campbell, E. B., Brockman, C. J., ... & Baade, L. E. (2012). Efficacy of Test of Memory Malinger Trial 1, Trial 2, the Retention Trial, and the Albany Consistency Index in a criterion group forensic neuropsychological sample. *Archives of Clinical Neuropsychology*, 28(1), 21-29.
- Skinner, N. F. (1999). When the going gets tough, the tough get going: Effects of order of item difficulty on multiple-choice test performance. *North American Journal of Psychology*, 1(1), 79-82.
- Stern, R. A., & White, T. (2003). *NAB, neuropsychological assessment battery: administration, scoring, and interpretation manual*. Psychological Assessment Resources.
- Strutt, A. M., Scott, B. M., Shrestha, S., & York, M. K. (2011). The Rey 15-item memory test and Spanish-speaking older adults. *The Clinical Neuropsychologist*, 25(7), 1253-1265.
- Tan, J. E., Slick, D. J., Strauss, E. & Hultsch, D. F. (2002). How'd they do it?

- Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, 16(4), 495-505.
- Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. Toronto, Ont.: Multi-Health Systems.
- Van Oorsouw, K. & Merckelbach, H. (2010). Detecting malingered memory problems in the civil and criminal arena. *Legal and Criminological Psychology*, 15(1), 97-114.
- Vyas, R., & Supe, A. (2008). Multiple choice questions: a literature review on the optimal number of options. *The National Medical Journal of India*, 21(3), 130-133.
- Wisdom, N. M., Brown, W. L., Chen, D., K., & Collins, R. L. (2012). The use of all three Test of Memory Malingering trials in establishing the level of effort. *Archives of Clinical Neuropsychology*, 27, 208-212.
- Wygant, D. B., & Lareau, C. R. (2015). Civil and criminal forensic psychological assessment: Similarities and unique challenges. *Psychological Injury and Law*, 8(1), 11-26.
- Young, S., Jacobson, R., Einzig, S., Gray, K., & Gudjonsson, G. H. (2016). Can we recognise malingerers? The association between malingering, personality traits and clinical impression among complainants in civil compensation cases. *Personality and Individual Differences*, 98, 235-238.