1-1-2016

Detection Of Malingering In Bona Fide Traumatic Brain Injury And Simulated Traumatic Brain Injury: Combining Response Time With Pvt Accuracy Results

Robert John Kanser
Wayne State University,

DETECTION OF MALINGERING IN BONA FIDE TRAUMATIC BRAIN INJURY AND SIMULATED TRAUMATIC BRAIN INJURY: COMBINING RESPONSE TIME WITH PVT ACCURACY RESULTS

by

ROBERT J. KANSER

THESIS

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

MASTER OF ARTS

2016

MAJOR: PSYCHOLOGY (Clinical)

Approved by:

_____

Advisor Date

ACKNOWLEDGMENTS

I owe special thanks to several people who made completion of this project possible. I would like to express my deepest gratitude to my advisor, Dr. Lisa J. Rapport, for her support, guidance, and encouragement throughout this entire process. I would also like to thank my thesis committee members, Dr. Robin A. Hanks and Dr. John L. Woodard, for their time and advice. Lastly, thanks to my fellow lab members for their assistance in recruitment and data collection.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Traumatic brain injury (TBI) is a serious health concern that is associated with a variety of behavioral and cognitive deficits. The severity and duration of these deficits varies greatly and depends on a number of factors. Currently, there is an estimated 5.3 million people living with TBI related deficits in the United States alone (Langlois, Rutland-Brown, & Wald, 2006).

Following injury, the ability to diagnose TBI accurately, predict future outcomes, and provide the most useful rehabilitative treatments relies heavily on information obtained through neuropsychological assessments. The validity of information obtained through these assessments, however, is greatly influenced by the amount of effort put forth by the examinee. Effort can account for as much as 50% of the variability in neuropsychological test scores (Meyers, Axelrod & Reinsch-Boothby 2011; Green, Lees-Haley & Allen 2001). An examinee providing suboptimal effort could, therefore, receive test scores that indicate drastically greater deficits than they truly have. Consequently, suboptimal effort negatively impacts the validity and utility of test results.

Feigned impairment or purposeful suboptimal effort associated with TBI is common, especially in compensation and litigation settings. Evidence suggests that base rates of suboptimal effort in this context approach 40% (Larrabee, Millis, & Meyers, 2009). As such, a great deal of research has been directed at creating standardized tests of effort to detect feigned impairment. Although advancements have been made, the accuracy of these tests in distinguishing bona fide TBI from feigned impairment is unacceptable. Moreover, many of these tests are vulnerable to coaching (Gunstad & Suhr, 2001; Rose, Hall, & Szalda-Petree, 1998) and information that threatens their security is readily available on the internet (Bauer & McCaffrey

2006). The aim of the proposed study is to determine the extent to which adding a covert measure to established tests of effort will improve diagnostic accuracy. Specifically, this study will examine the extent to which analysis of response time (latency) on a computerized version of the Test of Memory Malingering (TOMM-C; Tombaugh, 1996) improves its diagnostic accuracy in distinguishing between healthy adults providing full effort, TBI simulators, and individuals with bona fide TBI.

Section 1.1- Malingering

The validity of psychological assessments is contingent on the assumption that examinees provide full effort. The term *malingering* has been used to describe one type of suboptimal effort. The Diagnostic and Statistical Manual of Mental Disorders- Fifth Edition (DSM-5) defines malingering as "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs" (American Psychiatric Association, 2013, p. 726). There are two key features of this definition. First, the presentation of symptoms is conscious or "intentional." Second, these symptoms are presented in the context of an identifiable external incentive (i.e. material gain, avoiding punishment/formal responsibilities). Both of these concepts help differentiate malingering from other disorders in which inaccurate symptom presentation is common, namely factitious disorder and conversion disorder. Inaccurate symptom presentation in *factitious disorder* is thought to be volitional, or under conscious control, whereas the incentive is thought to be internal/psychological (i.e., play the sick role, receive attention). Symptom presentation in

conversion disorder is thought to be unconscious, and incentive also is considered internal/psychological (i.e. manage stress/conflict; Slick, Sherman & Iverson, 1999).

Although this definition provides a framework for conceptualizing malingering and distinguishes it from other disorders, its clinical utility is limited because the DSM-5 provides no concrete criteria for identifying and labeling malingering. Malingering is located in the V-Code section of the DSM-5 (V65.2). It is not classified as a mental disorder, but rather a behavior worthy of clinical attention. Without formal diagnostic criteria, the identification of malingering would rely almost entirely on clinical judgment. Although clinical judgment is critical to the assessment process, research has shown that it is vulnerable to individual biases and heuristics (mental shortcuts; Millis 2009). In an effort to improve identification, classification, and communication, researchers have offered their own definitions of malingering that include diagnostic criteria (Greiffenstein, Gola, & Baker, 1995; Rogers, 1990).

Slick, Sherman, and Iverson (1999) developed a definition and set of diagnostic criteria for malingering specific to neurocognitive dysfunction. Today, it is the most commonly used diagnostic system for assessing malingering in neuropsychological settings. Slick et. al. define malingered neurocognitive deficit (MND) as "the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility" (p. 552). Diagnosis of MND is a multi-method, multi-dimensional approach that requires the integration of data from self-reported symptoms, medical histories, behavioral observations, and neuropsychological testing.

Slick et. al. note that even with explicit, reliable criteria that integrate all possible sources of evidence, there remains uncertainty when inferring a client's volition, or conscious intent. Accordingly, their system provides three levels of classification corresponding to the degree of

diagnostic certainty: definite, probable, and possible malingering. Two criteria are common to all three of these diagnostic levels: a substantial external incentive must be present (e.g., compensation and pension, personal injury litigation) and the client's behavior must not be fully explained by psychiatric, neurological, or developmental factors. Definite, probable, and possible malingering are differentiated by the type (i.e., test data vs. self report) and the amount of evidence that indicates a *volitional exaggeration/fabrication* of symptoms.

Base rates of malingering vary depending on the context of the assessment (e.g., civil, criminal, medical) and the diagnosis from which the deficits are claimed to arise (Mittenberg, Patton, Canyock & Condit, 2002). A survey of members of the American Board of Clinical Neuropsychology revealed a staggering result: Approximately 30% of civil cases, 20% of criminal cases, and 8% of medical cases involved probable malingering (Mittenberg et al., 2002). Moreover, base rates of probable malingering are estimated to be 40-50% when external incentives are present (Larrabee, Millis & Meyers, 2009). With such high prevalence, it is clear that the ability to assess malingering accurately is critical, especially in areas where external incentives are common.

Section 1.2-Clinical Significance

Traumatic brain injury is the leading cause of disability in individuals under the age of 40, with an estimated 1.7 million individuals sustaining a TBI in the United States each year (Draper & Ponsford, 2008; Faul, Xu, Wald, & Coronado, 2010). Since 1991, the number of TBI related hospitalizations and visits to emergency rooms has steadily increased (Coronado, McGuire, Sarmiento, et al., 2012; Faul et al., 2010). This increase does not likely reflect a true increase in occurrence, but rather increased public knowledge and awareness of TBI and its

associated deficits. TBI has received significant media coverage, as it has been a major health concern for professional sports and veterans returning home from Iraq and Afghanistan (Coronado et al., 2012). With such prevalence and growing public awareness, it is not surprising that TBI-related cases are among the most common referrals in forensic neuropsychology (Larrabee, 2005; Ruff & Richardson, 1999). The forensic setting provides a variety of potential external incentives that increase the likelihood of malingering. Most notably, the potential monetary gain in TBI litigation is tremendous, with median rewards of $271,350 for mild TBI and $1,375,000 for moderate TBI (Kaiman, 2003). Multiple studies and reviews of the literature have converged to estimate the prevalence of malingering in mild TBI compensation cases and found it to be approximately 40% (see Larrabee, 2009, 2011). Clearly the risk and prevalence of malingering in the context of forensic traumatic brain injury assessment is great.

Beyond the forensic setting, TBI accounts for significant medical and rehabilitation costs. The annual direct medical costs of TBI are estimated to range between $9.1 billion and $14.6 billion in the United States alone (Finkelstein, Corso & Miller, 2006; Orman, Kraus, Zaloshnja & Miller, 2011). When accounting for indirect costs, such as loss of productivity, estimates exceed $76.5 billion annually (Finkelstein, Corso & Miller, 2006). The appropriate allocation of the medical and rehabilitation resources that contribute to these costs is contingent on the ability to diagnose TBI accurately; however, this cannot be done without assessing the amount of effort put forth during testing and the potential risk for malingering. Clearly, the inability to accurately distinguish between bona fide TBI and feigned cognitive impairment has drastic economic and social consequences for patients, the healthcare system, and the legal system.

Section 1.3-Assessment of Effort

A great deal of research has been directed at developing methods to detect suboptimal effort. Assessment of effort in the context of TBI is a particular focus of research due to its prevalence in settings where the risk of malingering is high and the fact that the majority of effort tests were developed using TBI samples (Millis, 2008).

The assessment of effort requires the integration and interpretation of information from a number of various methods. One fundamental approach is to look at *discrepancies between reliable sources of information*. For example, one set of the behavioral criteria in the Slick et al. classification system for MND pertain to discrepancies between test data and/or self report data and: observed behavior, reliable collateral reports, and documented medical history. This type of qualitative discrepancy analyses, however, relies heavily on clinical judgment. Studies have consistently shown that even expert clinicians are unable to identify suboptimal effort accurately using behavioral observations and test data alone (Ekman, O'Sullivan, & Frank 1999; Faust, 1995). These findings signaled the need for a more quantitative measure of effort, and led to the developmental of a number of tests designed to assess suboptimal effort.

Tests of effort have been called many things (e.g., malingering tests, tests of response bias, symptom validity tests, etc.). Larrabee (2012) has recommended the term *performance validity test* (PVT) for tests assessing effort, as it is more descriptive and makes no inferences regarding the examinees' volition. Stand-alone PVTs are tests created for the sole purpose of assessing suboptimal effort. They are the most frequently used, extensively studied, and best validated single measures of suboptimal effort (Constantinou, et al., 2005; Millis, 2008). Accordingly, their usage in nearly all neuropsychological evaluations has been deemed "medically necessary" by the National Academy of Neuropsychology (Bush et al., 2005) and the American Academy of Clinical Neuropsychology (Heilbronner et al., 2009). Although a number

of different PVTs used in cognitive assessment have been created, many share common features.

First, most stand-alone PVTs used in TBI assessment tap aspects of memory performance. Memory is a frequent target of symptom dissimulation during testing, as memory deficits are a common and well-known symptom of a wide variety of disorders (Binder & Rohling 1996; Suhr & Barrash, 2007). Over 80% of the general public is aware that a brain injury can result in memory deficits (Gouvier, Pretholdt, & Warner, 1988). Further, feigned memory impairment is among the most common strategies used by individuals instructed to simulate TBI (Bashem et al., 2014; Iverson 1995). Many PVTs also share a common structure: A simple target stimulus (i.e., line drawing, number, or symbol) is presented followed by a forced-choice recognition task in which the target stimulus is paired with a foil. Individuals must correctly identify the previously seen stimulus. This structure enables clinicians to detect a negative response bias. For example, if an individual performs significantly below chance, it is concluded that they were purposefully choosing incorrect items, or malingering.

Although below chance responding provides strong evidence for suboptimal effort, individuals suspected of malingering or those asked to simulate TBI rarely perform below chance on PVTs (Millis, 2008). They do, however, commonly perform significantly below healthy controls and individuals with bona fide TBI (Tombaugh, 1997). The majority of PVTs are tasks that are easy enough for individuals with neurocognitive deficits to respond correctly to nearly all items. Consequently, they utilize a concept known as “the floor effect,” or empirically-derived cut off scores that are well above chance (Bender & Rogers 2002; Neudecker & Skeel 2008). In doing so, they are able to increase sensitivity (the proportion of individuals providing suboptimal effort correctly identified as such by the test) while maintaining a clinically

acceptable specificity (the proportion of individuals providing optimal effort correctly identified as such).

The Test of Memory Malingering (TOMM; Tombaugh, 1996) is a 50-item forced-choice PVT that requires individuals to identify simple line drawings of common objects. The TOMM is a relatively easy task that has a cutoff score well above chance performance. It is one of the most commonly used and highly regarded measures of suboptimal effort among neuropsychologists (Sharland & Gfeller, 2007; Slick et al. 2004). Research has shown the TOMM is robust to differences in age, education, TBI, dementia, anxiety, and depression (Constantinou & McCaffrey, 2003; Tombaugh, 1996; Ashendorf, Constantinou, & McCaffrey, 2004; Rees, Tombaugh, & Boulay, 2001). Further, multiple studies have found the TOMM to have high specificity (Tombaugh, 1997; Rees, Tombaugh, Gansler, & Moczynski, 1998). Research on the sensitivity of the TOMM is far more variable. For example, studies with a known-groups design have used Slick et al. criteria to define MND and obtained sensitivity of 40-50% (Greve, Bianchini, & Doane 2006; Greve, Ord, Curtis, Bianchini, & Brennan, 2008). Conversely, several analogue studies in which college students were instructed to simulate TBI have shown sensitivities above 85% (Rees et al. 1998; Powell, Gfeller, Hendricks, & Sharland, 2004). These findings are similar to other well-established and validated PVTs, in which specificity tends to be high and sensitivity is moderate (Bianchini, Mathias, & Greve, 2010).

There are several potential explanations for the moderate sensitivity observed across most PVTs. First, most PVTs use cutoff scores that maximize specificity at the expense of reduced sensitivity (Bianchini et al., 2010). In clinical contexts, specificity is given precedence over sensitivity, as inaccurately labeling someone as malingering and denying them due resources is considered far more harmful than providing a true malingerer with undue resources. Next,

research has shown that PVTs are highly susceptible to coaching (Suhr & Gunstad 2007), and information that jeopardizes their security is readily available on the internet (Bauer & McCaffrey 2006). Moreover, a number of studies have shown that attorneys are likely to coach litigating clients prior to neuropsychological assessments. Surveys of practicing attorneys found that almost 50% believe they should provide specific information about tests (including validity measures) to their clients (Wetter & Corrigan 1995), and they will typically spend up to an hour discussing test content, detection of malingering, and common brain injury symptoms (Essig, Mittenberg, Peterson, Strauman, & Cooper 2001). Lastly, evidence suggests that PVTs have fairly high face validity. Tan, Slick, Strauss, and Hultsch (2002), found that fewer than 10% of participants asked to simulate TBI considered the TOMM to be a test of cognition, correctly identifying it as a measure of effort. As such, an increasing amount of research has been directed at developing new, covert measures of suboptimal effort.

Covert or 'embedded' measures of effort are scores or indices derived from standard cognitive tests. As such, these *embedded PVTs* may be less easily identified as measures of effort, and therefore less susceptible to coaching. Moreover, they provide useful information concerning both cognitive ability and effort without increasing the time required for testing. Although embedded PVTs are less robust to cognitive impairment than stand-alone PVTs (Miller, et. al., 2011), they have been shown to improve the diagnostic accuracy of suboptimal effort when used in combination with other PVTs (Larrabee, 2008). Consequently, a variety of embedded PVTs have been developed, and their usage is common in neuropsychological assessments (Meyers, et. al., 2011). Some of the most commonly used and extensively studied embedded PVTs include the Reliable Digit Span index derived from the Digit Span subtest of the Wechsler Adult Intelligence Scale 4th Edition (WAIS-IV; Greiffenstein, Baker, & Gola, 1994;

Miele et. al., 2012), and indices created for recognition and forced-choice trials of list-learning tasks such as the Rey Auditory Verbal Learning Test (AVLT; Meyers & Volbrecht, 2003) and the California Verbal Learning Test (CVLT; Coleman, Rapport, Millis, Ricker, & Farchione, 1998; Wolfe, Millis, Hanks, Fichtenberg, Larrabee, & Sweet, 2010).

Response time (RT) has been identified as a promising covert measure to distinguish between honest and feigned performance. Evidence suggests that slowed responding is one of the most common techniques deliberately employed by individuals instructed to simulate brain injury (Tan, Slick, Strauss, & Hultsch, 2002). Moreover, RT has been shown to be more resistant to coaching than performance accuracy (Rose et al., 1995, 1998). Therefore, several studies have looked at combining RT with conventional PVTs to improve their specificity and sensitivity. These studies have established that TBI simulators have longer average response times (Bolan, Foster, Schmand, & Bolan, 2002; Rees et al., 1998) with increased variability (Willison & Tombaugh 2006; Reicker 2008; van Hooff, Sargeant, Foster, & Schmand, 2007) compared to healthy controls putting forth full effort.

Section 1.4-Limitations of the Extant Literature

The most common research design in the study of malingering or performance validity is the analogue design (Suhr & Gunstad, 2007; Bianchini et al. 2010). In the analogue design, healthy adults are assigned to one of two groups: those instructed to perform to the best of their ability, and those instructed to feign TBI (sometimes being coached on how to do so). Although analogue design affords researchers great experimental control, its ecological validity has been questioned (Suhr & Gunstad, 2007). Characteristics found to differentiate healthy adults providing full effort and TBI simulators will not necessarily generalize to distinguish individuals

with bona fide TBI from those feigning impairment. Known-group designs, which are less common than the analogue design, use verified clinical groups and therefore have greater ecological validity. However, known-group designs typically do not include verified malingerers because few patients admit their status, and studies that include verified TBI groups tend to be limited by relatively small sample sizes (Greve et al., 2008). Thus, the strongest design would combine the strengths of analog and known-groups designs.

Unfortunately, very few studies have compared patterns of RT on PVTs among individuals with verified TBI and those simulating TBI. Of those studies that exist, investigations have generally been limited to simple comparisons of average RTs, neglecting the potential for complex patterns and analytic strategies. Additionally, mixed results have been found regarding which group (bona fide TBI or simulators) displays longer average RTs (Willison & Tombaugh 2006; Rose et al., 1995). In order to determine the clinical utility of RTs in improving the detection of suboptimal effort, additional research comparing individuals with bona fide TBI and those simulating TBI is necessary.

To date, only one study has combined RT data with a computerized version of the TOMM (TOMM-C) to distinguish controls, TBI simulators, and individuals with bona fide TBI (Vagnini, Berry, Clark & Jiang 2008). The results from this study, however, are constrained by a small sample size and a potential methodological flaw. The extreme variation in RT for simulators' correct trials on the TOMM-C reported is inconsistent with previous studies and indicates potential outliers. As such, the current study seeks to examine differences in response time for correct and incorrect responses across trials of the TOMM-C among healthy controls, TBI simulators, and individuals with bona fide TBI. These data will provide information

regarding the incremental utility of combining RT with one of the most commonly used PVTs, the TOMM.

Section 1.5-Aims of the current study

The current study sought to add to the literature concerning the use of RT to distinguish individuals with verified TBI from those instructed to feign memory impairment. The primary objective of the proposed study was to use information obtained from RT data to improve the diagnostic accuracy of a computerized version of one of the most well established PVTs, the TOMM-C. The main hypothesis was that individuals with bona fide TBI would display unique patterns of RT across trials of the TOMM-C, and that analysis of these patterns would improve the TOMM-C's diagnostic accuracy. This hypothesis was tested through the completion of two key objectives.

Objective 1*: Compare patterns of response times between full-effort healthy controls, individuals with bona fide TBI, and TBI simulators across correct and incorrect trials of the TOMM-C.*

Hypothesis 1a. It was predicted that average response times would be significantly longer for TBI simulators compared to individuals with bona fide TBI and full-effort healthy controls.

Hypothesis 1b. It was predicted that average variability in response time for correct and total responses would be greater for TBI simulators compared to individuals with bona fide TBI and full-effort healthy controls.

Hypothesis 1c. Response time analysis would reveal a distinct pattern across trials of the TOMM-C that reliably differentiates TBI simulators and individuals with bona fide TBI.

Objective 2*: Determine the extent to which RT characteristics provide incremental utility to the diagnostic accuracy of the TOMM-C.*

Hypothesis 2a. Predictive models that combine RT data with TOMM-C accuracy results would successfully distinguish controls, TBI simulators, and individuals with bona fide TBI.

Hypothesis 2b. The diagnostic accuracy of the TOMM-C would be improved by combining RT data with standard accuracy scores.

CHAPTER 2

METHOD

Section 2.1-Participants

Participants were 151 adults (96 men, 52 women) in three groups: *TBI Group*, *Healthy Comparison Group* (HC), and *TBI Simulators* (SIM). The TBI Group included 45 adults recruited from the Southeastern Michigan TBI Model System (SEMTBIS). All participants in the TBI group had a history of moderate to severe TBI indicated by: post-traumatic amnesia ≥ 24 hours, loss of consciousness ≥ 30 minutes, and Glasgow Coma Scale (GCS) < 13 at emergency department admission or abnormal neuroimaging. GCS at the time of admission to the emergency department ranged from 3 to 12 (M = 7.3, SD = 2.8). Participants in the SEMTBIS all sustained injuries severe enough to warrant inpatient rehabilitation treatment, were > 16 years old at the time of injury, and used English as their primary language. Additionally, participants in the TBI Group were at least 1 year post injury and able to participate in a valid assessment (e.g., sufficient attention capacity).

Neurologically healthy adults were recruited from the Detroit metropolitan area (n = 106). Inclusion criteria for these adults included English as their primary language and no history of neurological conditions. Forty-five healthy adults were assigned to the TBI Simulator Group. Sixty-one adults were assigned to the Healthy Comparison Group.

Age of participants ranged from 18 to 78 years. The HC (M = 45.7, SD = 16.8), SIM (M = 43.6, SD = 16.4), and TBI (M = 45.6, SD = 12.8) groups were equivalent on years of age, F (2, 148) = 0.25, p = .78. Education ranged from 18 to 20 years. The HC (M =13.9, SD = 2.4) and SIM (M = 14.6, SD = 2.0) did not differ significantly with respect to years of education;

however, both groups had more years of education than the TBI (M = 12.3, SD = 2.3) group, F (2, 148) = 12.57 p < .001.

Section 2.2- Measures

Test of Memory Malingering (TOMM; Tombaugh 1996). The TOMM is a 50-item visual recognition test designed to assess effort. The TOMM consists of two learning trials in which individuals view 50 consecutive line drawings of common objects for approximately 3 seconds each. The order of presentation is different between the two trials, but both are followed by a forced-choice recognition task in which the target item is paired with a foil. Individuals must correctly identify the previously seen item. The number of correct responses for each trial is tallied and can be compared to two cutoffs: below chance or criteria based on the performance of head injured and cognitively impaired individuals. According to the test manual, a raw score below 45 on Trial 2 suggests insufficient effort.

Section 2.3-Procedure

Section 2.3.1-Recruitment

The TBI Group (n = 45) was recruited from the pool of participants in the SEMTBIS who indicated willingness to be contacted for additional research opportunities. Participants in the SEMTBIS were pre-screened for suitability to participate and capacity to consent. The SEMTBIS provided data on injury severity as assessed via the Glasgow Coma Scale at the time of injury admission to the Emergency Department. The TBI participants were instructed to put forth full effort on all measures administered.

Neurologically healthy participants (n = 106) were recruited from the Detroit

Metropolitan area through newspaper advertisements and flyers posted around the campus of Wayne State University. Potential participants were screened over the telephone to determine their eligibility. Participants recruited for the neurologically healthy groups were excluded from the study if they reported a history of neurological conditions (e.g., Alzheimer's disease, seizure disorder, etc.) or a history of TBI. The *Healthy Comparison Group* (n = 61) was instructed to put forth full effort on all measures administered.

The *TBI simulator group* (SIM, n = 45) participants were read a scenario that describes their involvement in litigation for a TBI they sustained following a motor vehicle accident. The script from this scenario has been used successfully in TBI simulation studies with similar research designs (DenBoer & Hall, 2007; Tombaugh, 1997). SIM participants were then given time to read a pamphlet that describes common symptoms that can occur following TBI (Coleman, Rapport, Millis, Ricker, & Farchione, 1998; Rapport, Farchione, Coleman, & Axelrod, 1998). After the induction procedure SIM group participants completed the remainder of the assessment battery under instructions to feign TBI.

Informed consent procedures were completed with all participants in accordance with the institutional review board guidelines. All testing took place at the research laboratory of the primary investigator and the Rehabilitation Institute of Michigan. Testing for each participant was completed in a single session lasting approximately 2 hours. All study participants were compensated $30.

Section 2.3.2- Debriefing

Following completion of the battery, all SIM participants were administered a 6-item questionnaire. Questions included whether they tried to simulate TBI as instructed, their

strategies to do so, and how difficult they rate the experience of simulating TBI. SIM participants were excluded from analysis if they reported not attempting to simulate TBI.

Section 2.4-Statistical Analyses

Prior to analysis, the data were screened according to recommendations by Tabachnick and Fidell (2012), including assumptions of parametric model (e.g., skewness, winsorizing outliers > 3 z, homogeneity of variance, and collinearity). Per standard protocol for RT data, responses < 250ms were considered invalid because it is faster than could be cognitively processed. Fortunately, there were very few invalid data points of this nature. Of the few cases that had these invalid data points, none had more than two in either of the 50-item trials of the TOMM. *Descriptive statistics* were conducted to describe the sample demographics and TOMM-C performance. To establish that the groups are demographically equivalent, they were compared across age and years of education using one-way analysis of variance (ANOVA).

Hypotheses 1a and 1b were tested using a mixed-design ANOVA, with Group (HC, SIM, TBI) as the between-subjects factor, and TOMM Trial (1, 2) and index (Mean RT, CV variability) as within-subject factors. For this analysis, the variables Mean RT and Mean CV were converted to a common metric (z) in order to compare within-group profile of RT performance. Further analyses incorporated ANOVA and Kruskal-Wallis tests, as appropriate (for tests with severely skewed distributions), with group (TBI, HC, SIM) as the between-subjects factor. Per the hypotheses, these tests assessed group differences on average RT (Hypothesis 1a) and RT variability (Hypothesis 1b). These group comparisons were performed for both Trial 1 and Trial 2 of the TOMM-C. Post hoc comparisons were conducted as appropriate, using LSD tests ($p < .05$).

Hypothesis 1c sought to extend RT analysis by identifying novel patterns of RT capable of distinguishing TBI simulators from individuals with bona fide TBI. Analyses included comparisons of frequencies of lengthy RTs for responses (TOMM 1&2 Lengthy), ratios of average RT for incorrect to correct trials (TOMM 1&2 RT C/I), and differences in average RT for correct and incorrect trials (TOMM 1&2 RT Difference).

Classification accuracy statistics (hit rate, sensitivity, specificity) examined the diagnostic validity of the TOMM-C accuracy and RT indices. Negative Predictive Powers (NPP) and Positive Predictive Powers (PPP) were calculated at base rates of clinical relevance (40% and 10%). Hypothesis 2a and 2b were tested using multivariable binary logistic regressions, testing the individual and combined predictive values of the traditional accuracy index and various indices derived from RT data. Logistic regression models with group membership (TBI vs. SIM and SIM vs. HC) as the outcome variable were fitted for each RT index separately. Nagelkerke R^2 values were generated to evaluate the variance accounted for by individual and combined indices. Multivariable logistic regression models combined TOMM-C accuracy and RT variables as covariates, with group membership (TBI vs. SIM and SIM vs. HC) as the outcome variables. The combined models were examined to determine the extent to which RT data could improve model classification over standard scoring using TOMM2 accuracy (number correct) through analysis of Bayesian information criterion (BIC) and receiver operating characteristics (ROC) curve analysis.

CHAPTER 3

RESULTS

Table 1 presents descriptive statistics and group comparisons of TOMM-C indices. To conduct the mixed-design ANOVA, which tested within-group profile of RT performance, the variables Mean RT and Mean CV were converted to a common metric (z). The analysis tested Group (HC, SIM, TBI) as the between-subjects factor, and TOMM Trial (1, 2) and index (Mean RT, CV variability) as within-subject factors. The results of the mixed-design ANOVA revealed a main effect of group and a group x index interaction. The main effect of group indicated that across both average time and variability, SIM scored significantly higher than TBI, who scored significantly higher than HC, $F(1, 148) = 33.89$, $p < .001$, $\eta^2 = .31$. Figure 1 depicts the Group x Index interaction, $F(1, 148) = 11.44$, $p < .001$, $\eta^2 = .13$. No other main effects or interactions (TOMM Trial x Group Membership, TOMM Trial x Index, or TOMM Trial x Index x Group Membership) were significant (ps > .141).

Table 1 also presents univariate group comparisons of TOMM-C indices. ANOVAs indicated that a number of TOMM-C indices differed significantly across group. TOMM-C accuracy and RT indices differing significantly across groups with large effect sizes ($\eta^2 > .26$) included: TOMM1 Correct, TOMM2 Correct, TOMM1 RT correct, TOMM2 RT correct, TOMM1 RT mean, and TOMM2 RT mean. Of the RT indices, TOMM 2 RT mean had the largest effect size ($F(2, 148) = 43.98$, $p < .001$, $\eta^2 = .37$). Post hoc analyses (LSD tests) revealed that nearly all of the RT indices differed significantly between SIM and HC. All of the RT indices were significantly larger for SIM compared to HC, with the exception of TOMM1 RT difference, TOMM2 RT difference, and TOMM2 RT C/I. Slightly fewer RT indices differed

significantly between SIM and TBI; however, all of the RT indices that showed significant group differences were larger for SIM compared to TBI.

Table 2 presents descriptive correlations for TOMM-C indices and demographic variables. As would be expected, the component RT variables (e.g., TOMM1 RT Correct, TOMM1 RT Incorrect) were highly correlated with the average RT for total items of the corresponding trial (e.g., TOMM1 RT mean). Pearson correlations between component RT variables and their inclusive RT trials ranged from .66 to .99. Of note, the difference scores (e.g., TOMM1 RT difference, TOMM2 RT difference) and ratio scores (e.g., TOMM1 RT C/I, TOMM1 RT C/I) within corresponding trials were very highly correlated ($r \geq .90$), indicating that these indices were essentially redundant. Accordingly, the difference scores were dropped from further analyses. Also of note, neither age nor education showed meaningful correlations with any TOMM-C accuracy or RT variables (all but one correlation $r < .20$).

Overall Classification Accuracy of the TOMM-C

As expected, TOMM-C pass/fail classification dictated by the standard cutoff score was significantly related to group membership and displayed excellent group discrimination, (χ^2 2, N = 151) = 94.96, $p < .001$, $\varphi = .79$. None of the HC, 2.2% ($n = 1$) of the TBI, and 73.3% ($n = 33$) of the SIM failed the TOMM-C.

Logistic Regressions for Single-Variable Models (SIM vs. HC)

Table 3a presents classification accuracy statistics for the TOMM indices as individual predictors of group status (SIM vs. HC). TOMM2 Correct displayed the largest hit rate (91%) and specificity (98%). The largest sensitivity (100%) was observed when using RT indices that

incorporated TOMM Trial 2 errors (e.g., TOMM2 RT incorrect, TOMM2 RT C/I). However, this result is largely due to the fact that so few HC (n = 13) had at least one error on Trial 2 in comparison to SIM (n = 38) and were, therefore, all predicted to be SIM. Excluding these two RT indices, TOMM1 Correct displayed the largest sensitivity (87%). The RT indices performed modestly with respect classification accuracy. Of the RT indices, TOMM2 RT mean performed the best, with a hit rate of 82%, sensitivity of 73%, and specificity of 82%.

Table 3b provides statistics from the logistic regressions, including the chi-square statistics testing the significance (reliability) of the models, as well as the odds ratios and information on the significance of these models. The following single-variable models were significant (p < .05) predictors of group membership: TOMM1 Correct, TOMM2 Correct, TOMM1 RT correct, TOMM2 RT correct, TOMM1 RT mean, TOMMM2 RT mean, TOMM2 RT CV, and TOMM1 RT C/I. In order to quantify the discriminability of these models, area under the receiver operating characteristics (ROC) curve values were calculated (see Table 3a). Area under the curve (AUC) range from .50 to 1.0, with larger values indicating better discrimination. AUC values can be considered "acceptable" (.70 ≤ AUC ≤ .79), "excellent" (.80 ≤ AUC ≤ .89), or "outstanding" (AUC ≥ .90; Hosmer & Lemeshow, 2000). The following variables showed "outstanding" discriminability: TOMM1 Correct, TOMM2 Correct, TOMM2 RT correct, and TOMM2 RT mean. Discrimination was "excellent" with TOMM1 RT correct and TOMM1 RT mean and "acceptable" with TOMM 1 RT C/I and TOMM2 RT CV. Discrimination with the remaining models (TOMM1 RT incorrect, TOMM2 RT incorrect, TOMM1 RT CV, and TOMM2 RT C/I) was unacceptable (AUC < .70; range .56 to .65). The remaining significant models all displayed "acceptable" or "excellent" discriminability. TOMM1 RT incorrect, TOMM2 RT incorrect, TOMM1 RT CV, and TOMM2 RT C/I each showed AUC

< .70 “acceptable” (range .56 to .65).

Classification Accuracy of Two-Variable Models (SIM vs. HC)

Tables 3a and 3b show classification and model fit statistics for the multivariable logistic regression models predicting group membership (SIM or HC). Each of the two-variable models combined TOMM2 Correct with one of the RT indices to determine the extent to which the RT index could add incremental predictive value to TOMM2 Correct (continuous score) in predicting group membership (SIM or HC). Table 3b shows that all of the two-variable models were significant ($p < .001$); however, RT indices that added *incremental* predictive value ($p < .05$) to TOMM2 Correct were: TOMM1 RT correct, TOMM1 RT incorrect, TOMM2 RT correct, TOMM1 RT mean, and TOMM2 RT mean. These combined models led to increased AUC values that ranged from .94 to .99 (i.e., increments in AUC of .04 to .09 as compared to TOMM2 Correct AUC = .90). It is important to note, however, that AUC values are fairly insensitive to changes in model fit when multiple covariates are used within the same model. Accordingly, Bayesian information criterion (BIC) statistics were calculated. The BIC statistic is used to quantify a covariate’s incremental predictability to the model while favoring model parsimony. In other words, an added covariate must contribute enough incremental predictability in order to overcome the “penalty” for increasing the number of covariates and be “preferred.” BIC statistics are interpreted by comparing differences in BIC values across models. Model preference is then classified as weak, positive, strong, or very strong (Raftery, 1996), with negative values being desirable. According to Raftery, 1996), an absolute BIC difference of 0 – 2 is considered a “weak preference,” 2 – 8 is considered a “positive preference,” 8 – 10 is considered “strong” preference, and a difference greater than 10 is considered “very strong” preference for the model.

As compared to BIC for the single-variable model with TOMM2 Correct (-427.54), four two-variable models were "very strongly" preferred: TOMM1 RT Correct, TOMM2 RT Correct, TOMM1 RT mean and TOMM2 RT mean. The remaining two-variable models were not preferred over the single-variable model for TOMM2 Correct.

Classification Accuracy of Single-Variable Models (SIM vs. TBI)

Table 4a presents classification accuracy statistics for the TOMM indices as individual predictors of group status (SIM vs. TBI). TOMM2 Correct displayed the largest hit rate (87%), sensitivity (78%), and specificity (96%). Once again, the large sensitivities observed in RT indices that incorporated errors were greatly influenced by the fact that so few TBI had at least one error. Of the other RT indices, TOMM2 RT mean had the largest hit rate (66%), sensitivity (53%), and specificity (78%).

Table 4b shows the single-variable models that were significant ($p < .05$) predictors of group membership. Indices based on TOMM continuous accuracy scores were significant predictors with TOMM1 Correct (AUC = .85) and TOMM2 Correct (AUC = .88) having excellent discriminability. A number of RT indices were significant predictors; however, only TOMM1 C/I displayed adequate discriminability (AUC = .70). Notably, TOMM2 RT mean was once again a significant predictor ($\chi^2 = 15.35$, $p < .001$, Nagelkerke $R^2 = .21$) and had an AUC very near acceptable discriminability (AUC = .69, 95% C.I. = 0.58-0.80).

Classification Accuracy of Two-Variable Models (SIM vs. TBI)

Tables 4a and 4b show classification and model fit statistics for the multivariable logistic regression models predicting group membership (SIM vs. HC). A number of the two-variable

models had “outstanding” discriminability (AUC > .9) that was larger than the single-variable TOMM2 Correct model (AUC = .88). The two-variable TOMM2-C + TOMM1 RT mean models had the largest AUC (.97). The two-variable TOMM2-C + TOMM1 RT mean and TOMM2-C + TOMM1 RT correct led to slight increases in hit rate, sensitivity, specificity, AUC, and Nagelkerke R^2 above the single-variable TOMM2 Correct model. However, the magnitude of the BIC difference indicated only a “weak” preference for the two-variable models that tested TOMM1 RT correct, TOMM2 RT correct, TOMM1 RT mean and TOMM2 RT mean. In sum, the gains in classification accuracy and model fit did not far exceed the cost associated with loss in parsimony.

CHAPTER 4

DISCUSSION

Findings provide support for the hypothesis that combining RT data with performance accuracy on the Test of Memory Malingering (TOMM) can improve its diagnostic accuracy. A number of models that combined novel and previously investigated RT indices with TOMM-C accuracy led to excellent diagnostic accuracy. Moreover, several of these RT indices added incremental predictive value that was preferred over using TOMM-C accuracy in isolation. The extent of this incremental predictive value and preference was greatly influenced by two factors. First, TOMM-C accuracy was an excellent individual predictor of group membership, making very few errors in classification. Accordingly, it was very difficult for new indices to add incremental predictive value. Second, preference for RT indices was greatly influenced by which groups were compared. There was a strong preference for combining RT indices with TOMM-C accuracy to distinguish healthy adults from individuals instructed to feign TBI. In contrast, there was weak preference for combining TOMM-C accuracy with RT indices to distinguish individuals with TBI from those feigning TBI. These findings highlight the importance of including individuals with bona fide TBIs when evaluating and developing performance validity measures.

Objective 1: *Compare patterns of response times across groups*

Consistent with previous research, findings show that individuals simulating TBI have longer RTs (Bolan, Foster, Schmand, & Bolan, 2002; Rees et al., 1998) with increased variability (Willison & Tombaugh 2006; Reicker 2008; van Hooff, Sargeant, Foster, & Schmand, 2007) compared to healthy controls putting forth full effort. Findings add to the limited and conflicted

literature concerning whether individuals feigning TBI or individuals with verified TBI display longer RTs (Willison & Tombaugh, 2006; Rose et al., 1995). Consistent with Hypothesis 1a, individuals feigning TBI had longer average RTs than individuals with verified TBI, who had longer average RTs than healthy adults. This difference in RT is likely due, in part, to simulators adopting the strategy of slowed responding and overestimating the tendency toward slowed cognitive processing among people with TBI. Slowed responding has been identified as one of the most common techniques deliberately employed by individuals instructed to simulate brain injury (Tan, Slick, Strauss, & Hultsch, 2002). Additionally, it is possible that this difference in RT is partly due to the additional cognitive processing simulation requires. For example, a person providing full effort must identify the correct response option and select it. In contrast, a person simulating TBI must not only identify the correct response option, but also decide whether to select the correct or incorrect response option. This decision requires additional time and can be further lengthened by other cognitive processes (i.e., trying to remember how many items one has answered incorrectly to that point). Taken together, these two explanations could account for longer RTs for simulators who adopted a strategy of slowed responding, and also those who did not.

Findings support Hypothesis 1b that variability in RTs would be different across the groups; however, this difference was only observed on Trial 2 of the TOMM-C. It is unclear why this pattern did not hold for Trial 1 of the TOMM-C. The pattern of findings indicates that simulators maintained relatively high variability across both trials of the task, whereas adults with TBI and healthy adults providing full effort showed relatively reduced variability on the second trial. It could be that variability among the three groups was equivalent during the first trial, while adjusting to a novel task (i.e., task instructions, method of responding, etc.). By Trial

2, however, participants had become accustomed to the task. Thus, in Trial 2, individuals putting forth full effort (i.e., verified TBI and healthy adults) commit fewer errors, and RT variability drops as a natural consequence of speed and certainty on correct trials relative to incorrect trials. In contrast, simulators override natural patterns of responding, and variability remains stable despite increased familiarity with the task. Of note, consistent with prior research, adults with TBI showed greater variability compared to healthy adults providing full effort.

Hypothesis 1c sought to extend the research literature by examining whether group differences existed with respect to a novel set of RT indices. Specifically, the current study investigated whether examination of RTs for incorrect trials could distinguish groups. Findings show that indices that utilize RTs for incorrect trials (e.g., ratio of RT for correct to incorrect trials, mean difference between incorrect and correct trials) do differ among the groups. Individuals providing full effort (e.g., healthy adults and individuals with TBI) had RTs for incorrect items that were longer than their RTs for correct items. In contrast, individuals simulating TBI had item RTs that were comparable, regardless of accuracy. This finding lends further support for the hypothesis that differences in cognitive processing contribute to RT differences. Individuals providing full effort may have longer RTs for incorrect than correct trials because there is likely greater uncertainty in the answer. In contrast, individuals feigning TBI have similar RTs for incorrect and correct trials because their processing is the same (e.g., identifying the correct answer and deciding whether to answer correctly). It is important to note that group differences on RT indices incorporating incorrect trials were not observed on Trial 2 of the TOMM. This finding is likely due to the fact that very few individuals providing full effort commit any errors on Trial 2, which translates to reducing the number of cases for the index; the remaining cases represent a select subgroup and reduce statistical power considerably. In sum,

comparing errors and correct trials seems somewhat promising; however, the TOMM-C yields such limited variability in performance accuracy that this area of study would be better investigated using a PVT that yields greater performance accuracy variability.

Objective 2: *Determine the incremental utility of combining RT indices with TOMM-C accuracy*

To date, one study has combined RT data with the TOMM-C with a similar three-group design (Vagnini, Berry, Clark & Jiang 2008). However, the results from this study were constrained by a very small sample and extreme variation in RT for one group, suggesting the presence of unaccounted outliers. Moreover, the TOMM-C classified their entire sample with 100% accuracy. As such, the study by Vagnini and colleagues was unable to assess the effect of combining RT indices with TOMM-C accuracy in distinguishing individuals feigning TBI from individuals with verified TBI. The central hypothesis of the present study was that combining RT indices with traditional TOMM-C accuracy (total correct) could enhance its diagnostic accuracy. Findings provide some support for this hypothesis. A number of models that combined novel and previously investigated RT indices with TOMM-C accuracy led to excellent diagnostic accuracy. Moreover, several of these RT indices added incremental predictive value that was preferred over using TOMM-C accuracy in isolation.

As expected, traditional accuracy scores for Trial 1 and 2 of the TOMM were the best single predictors of group membership, showing "excellent" (TBI vs. SIM) and "outstanding" (SIM vs. HC) group discrimination. Of the RT indices, average RT for Trial 1 and 2 of the TOMM-C were the best predictors of group membership, showing "excellent" (Trial 1) and "outstanding" (Trial 2) discrimination of individuals feigning TBI and healthy adults. However, all RT indices were less successful in discriminating individuals feigning TBI from those with

verified TBI. Only average RT for Trial 2 approached near "acceptable" discrimination for these groups.

The process of evaluating enhancements in diagnostic accuracy relied on investigating changes in hit rate, sensitivity, specificity, and AUC values. Unfortunately, there is no standard for weighting the importance of these statistics in evaluating changes in diagnostic accuracy. For the purposes of this study, AUC was ranked as the most important statistic, followed by hit rate, specificity, and sensitivity (Hosmer & Lemeshow, 2000). AUC was chosen as the most important statistic because it is an objective measure that considers the ranges of sensitivities and specificities in the sample.

Findings show that a number of the predictive models that combined TOMM-C accuracy with RT indices led to improved diagnostic accuracy over prediction using TOMM-C accuracy alone. However, only average RT for total and correct items on Trial 1 and 2 of the TOMM-C added incremental predictive value to TOMM-C accuracy with respect to group discrimination (TBI vs. SIM and HC vs. SIM). It is important to note that this incremental predictive value does come at the cost of using more than one measure to predict group membership. Statistics like the BIC take this cost into consideration and classify a predictive model's preference for added variables while favoring parsimony (Raftery, 1996).

The degree to which the incremental predictive value provided by the aforementioned RT indices was preferred by the model was greatly influenced by which groups were compared. There was a strong preference for using the RT indices when distinguishing healthy adults from individuals feigning TBI. In contrast, there was a weak preference for using the RT indices when distinguishing individuals feigning TBI from those with verified TBI. One could argue, however,

that RT data should not pay a heavy price for decreasing model parsimony. Parsimony is important when the added index would be costly in meaningful ways, such as increased time in testing or financial cost associated with adding an additional test to the assessment battery, or if calculating the new index would be labor intensive. Although the process of creating the computerized version programming the TOMM was moderately labor-intensive, adding RT indices in the clinical setting could be very low investment of resources, because it is inherent in the task, adding no time to the assessment battery and little added effort from clinicians. Regardless, it is interesting that RT adds more incremental predictive value to discriminating healthy adults from individuals feigning TBI than discriminating individuals feigning TBI from individuals with verified TBI. This finding suggests that RT indices are less robust to TBI than TOMM-C accuracy. Although individuals feigning TBI have longer average RTs than individuals with TBI, there is greater overlap in their RT distributions than their accuracy distributions. This makes sense given that slowed processing speed is one of the hallmark symptoms of TBI (Cicerone et al., 2011; Dikmen et al., 2009; Axelrod et al., 2001, 2002).

Limitations

The most prominent limitation of the present study is related to the sample of participants and its generalizability. In order to increase experimental control, the TBI group consisted of individuals with well-documented histories of moderate to severe TBI. Accordingly, the extent to which findings generalize to individuals with uncomplicated, mild or very severe TBI is unknown, and independent replication with these populations is necessary.

Additionally, the generalizability of the simulator group to individuals who feign TBI in clinical and forensic settings may be limited. This is a common limitation of analog designs has

been previously noted (Larrabee, 2007; Rogers, 1997; Suhr & Gunstad, 2007), with studies showing far greater TOMM sensitivity in analog (Rees et al. 1998; Powell, Gfeller, Hendricks, & Sharland, 2004) as compared to known-groups designs (Greve, Bianchini, & Doane 2006; Greve, Ord, Curtis, Bianchini, & Brennan, 2008). In contrast to the significant incentives for successfully feigning TBI in forensic settings, the simulator group in the present study did not have any external incentive for avoiding detection. Additionally, simulators did not have the time to prepare (i.e., utilize the information on the internet that threatens PVT test security; Bauer & McCaffrey 2006) for testing that an individual in a clinical or forensic setting would. Lastly, the method of coaching used in the present study has been common practice in analog designs. However, research shows that the coaching individuals receive from their attorneys is far more sophisticated and detailed (Wetter & Corrigan 1995; Essig et al., 2001). All of these factors may have contributed to less sophisticated and effective feigning strategies that enabled only 26.7% of simulators to pass the TOMM-C. It is important to note that TOMM-C sensitivity in the present study was comparable to those observed in other studies employing analog designs (Rees et al. 1998; Powell, Gfeller, Hendricks, & Sharland, 2004).

Another limitation is that our groups were not equivalent with respect to education. More specifically, the simulator and healthy adult comparison group had significantly more years of education than the TBI group. However, education was not meaningfully related to any of the TOMM-C RT indices. Moreover, because IQ and education facilitate ability to feign successfully (Rapport et al., 1998), one could argue that an advantage in years of education favored the simulator group, making it more likely for them to avoid detection using TOMM-C traditional accuracy scores. Even in this context of increased challenge to the TOMM, it performed well using traditional accuracy and via added RT indices.

Conclusions and Future Directions

This study contributes to the limited body of research examining the incremental utility of combining RT with traditional PVTs in distinguishing feigned and bona fide TBI. Findings validate previous RT research that has consistently shown individuals feigning TBI produce longer RTs than healthy adult comparisons (Bolan et al., 2002; Rees et al., 1998; Willison & Tombaugh 2006; Reicker 2008; van Hooff et al., 2007). It also provides some clarity to the limited, mixed findings surrounding whether or not individuals feigning TBI display longer RTs than individuals with verified TBI (Willison & Tombaugh 2006; Rose et al., 1995). Analyses comparing RTs for correct to incorrect items show promise as measures of performance validity, providing the task difficulty is increased such that a majority of examinees commit some errors. Future studies should investigate this avenue of research using PVTs that generate more accuracy variability.

The ability of RT indices to add incremental predictive value to TOMM-C accuracy was somewhat limited by the TOMM-C's excellent group discrimination. Future studies using analog designs should consider employing techniques that could lead to more sophisticated feigning strategies and, therefore, classification rates that are comparable to those in clinical/forensic settings. Despite the excellent classification accuracy of the TOMM-C through traditional scoring, RT indices provided incremental predictive value to group distinction. The degree of preference for these RT indices depended on what groups were being discriminated, revealing the importance of including a group with bona fide TBI when developing and adapting PVTs. Future research using a more sophisticated group of individuals feigning TBI and a group of individual with mild TBI may prove beneficial in further evaluating the clinical utility of

combining RT indices with TOMM-C.

Appendix A (Tables)

Table 1. *Descriptive Statistics and Group Comparisons of TOMM Performance for TBI (n = 45), HC (n = 61) and SIM (n = 45) Groups.*

	HC		*SIM*		*TBI*						
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*df*	*p*	η^2	Significant Contrasts
TOMM1 Correct	47.7	(2.8)	33.5	(7.9)	43.3	(5.0)	92.05	2, 148	< .001	.55	HC > TBI > SIM
TOMM2 Correct	49.7	(0.7)	35.8	(11.1)	49.2	(1.4)	79.52	2, 148	< .001	.52	HC = TBI > SIM
TOMM1 RT correct	1401	(335)	2445	(1010)	1913	(462)	34.43	2, 148	< .001	.32	SIM > TBI > HC
TOMM1 RT incorrect	2421	(1182)	2694	(1169)	2575	(998)	0.66	2, 125	.520	.01	--
TOMM2 RT correct	1109	(237)	2018	(792)	1513	(374)	42.71	2, 148	< .001	.37	SIM > TBI > HC
TOMM2 RT incorrect	2058	(1131)	2210	(1000)	1861	(743)	0.84	2, 67	.438	.02	--
TOMM1 RT mean	1441	(355)	2476	(983)	1989	(504)	34.02	2, 148	< .001	.32	SIM > TBI > HC
TOMM2 RT mean	1114	(246)	2017	(770)	1515	(364)	43.98	2, 148	< .001	.37	SIM > TBI > HC
TOMM1 RT CV	0.45	(0.19)	0.49	(0.16)	0.46	(0.16)	1.02	2, 148	.364	.01	--
TOMM2 RT CV	0.32	(0.14)	0.42	(0.12)	0.37	(0.10)	7.48	2, 148	.001	.09	SIM = TBI > HC
TOMM1 RT difference	979	(1142)	208	(670)	646	(705)	8.78	2, 125	< .001	.12	HC > TBI > SIM
TOMM2 RT difference	661	(819)	118	(467)	272	(701)	3.83	2, 67	.027	.10	HC > TBI = SIM
TOMM1 RT C/I	0.72	(0.35)	0.94	(0.20)	0.79	(0.20)	8.13	2, 125	< .001	.12	SIM > TBI = HC
TOMM2 RT C/I	0.33	(0.15)	0.33	(0.15)	0.33	(0.15)	2.59	2, 67	.083	.07	HC > TBI = SIM
TOMM1 Lengthy[1]	3.8	(4.0)	18.4	(14.3)	10.5	(8.3)	45.46	2	< .001	.30	SIM > TBI > HC
TOMM2 Lengthy[1]	2.2	(3.1)	18.9	(15.5)	8.3	(7.5)	62.92	2	< .001	.35	SIM > TBI > HC

Note. TOMM1 = Test of Memory Malingering–Trial 1; TOMM2 = Test of Memory Malingering–Trial 2; RT CV = Response Time coefficient of variation; RT difference = RT incorrect – RT correct; RT C/I = ratio correct / incorrect.
1. Kruskal-Wallis and Mann-Whitney tests.

Table 2. *Descriptive Correlations for TOMM Accuracy and Response Time (RT) Indices.*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. TOMM1 Correct	1.00														
2. TOMM2 Correct	.83**	1.00													
3. TOMM1 RT correct	-.52**	-.38**	1.00												
4. TOMM1 RT incorrect	-.01	.05	.57**	1.00											
5. TOMM2 RT correct	-.57**	-.45**	.86**	.41**	1.00										
6. TOMM2 RT incorrect	-.15	-.01	.66**	.52**	.71**	1.00									
7. TOMM1 RT mean	-.51**	-.34**	.99**	.66**	.84**	.66**	1.00								
8. TOMM2 RT mean	-.56**	-.42**	.86**	.41**	.99**	.76**	.83**	1.00							
9. TOMM1 RT CV	-.16*	-.18*	.16*	.34**	.06	-.18	.19**	.04	1.00						
10. TOMM2 RT CV	-.38**	-.27**	.28**	.01	.41**	.18	.28**	.41**	.23**	1.00					
11. TOMM1 RT difference	.46**	.41**	-.21**	.67**	-.26**	-.16	-.10	-.25**	.24**	-.19*	1.00				
12. TOMM2 RT difference	.23*	.32**	.00	.03	-.11	.62**	.01	-.03	-.28*	-.10	.08	1.00			
13. TOMM1 RT C/I	-.43**	-.39**	.25**	-.56**	.28**	.08	.15*	.27**	-.16*	.29**	-.90**	-.10	1.00		
14. TOMM2 RT C/I	-.14	-.20*	.03	-.03	.17	-.52**	.02	.09	.23*	.20*	-.11	-.91**	.13	1.00	
15. Age	.16*	.17*	-.02	.12	-.08	-.12	-.02	-.07	.12	-.09	.17*	.05	-.13	-.01	1.00
16. Education	-.14*	-.22**	.17*	.19*	.09	.13	.18*	.10	.13	.03	.06	.00	-.09	-.17	.14*

Note. TOMM1 = Test of Memory Malingering–Trial 1; TOMM2 = Test of Memory Malingering–Trial 2; RT CV = Response Time coefficient of variation; RT difference = RT correct – RT incorrect; RT C/I = ratio correct / incorrect.
*$p < .05$, **$p < .01$.

Table 3a. *Classification Statistics: TOMM Performance for Single and Two-variable Models Predicting Simulator (SIM) and Full Effort Healthy Comparison (HC) Group Membership*.

	Hit Rate	Sn	Sp	PPP BR 40%	NPP BR 40%	PPP BR 10%	NPP BR 10%	R^2	AUC	AUC 95% CI	BIC
One-Variable Models:											
1. TOMM1 Correct	.89	.87	.90	.86	.90	.47	.99	.79	.97	[.94, .99]	-434.76
2. TOMM2 Correct	.91	.80	.98	.97	.89	.80	.98	.75	.90	[.83, .98]	-427.54
3. TOMM1 RT correct	.78	.67	.87	.78	.80	.37	.97	.50	.83	[.75, .91]	-389.90
4. TOMM1 RT incorrect	.54	.53	.55	.44	.63	.12	.92	.02	.58	[.46, .70]	-268.36
5. TOMM2 RT correct	.81	.71	.89	.81	.83	.44	.97	.61	.90	[.83, .96]	-404.34
6. TOMM2 RT incorrect	.75	1.00	.00	.39	.00	.10	.00	.01	.56	[.36, .76]	-134.98
7. TOMM1 RT mean	.78	.69	.85	.74	.81	.33	.96	.50	.84	[.75, .92]	-390.55
8. TOMM2 RT mean	.82	.73	.82	.74	.83	.32	.96	.62	.91	[.85, .97]	-405.95
9. TOMM1 RT CV	.57	.11	.90	.45	.60	.09	.91	.02	.62	[.51, .72]	-342.28
10. TOMM2 RT CV	.63	.36	.84	.60	.66	.21	.92	.15	.77	[.68, .86]	-353.04
11. TOMM1 RT C/I	.71	.71	.70	.61	.79	.20	.95	.17	.74	[.64, .85]	-279.52
12. TOMM2 RT C/I	.86	1.00	.46	.54	1.0	.17	1.0	.18	.65	[.44, .86]	-141.31

Note. Sn = Sensitivity (detection of simulated TBI), Sp = Specificity (bona fide TBI); PPP = Positive Predictive Power, NPP = Negative Predictive Power (each presented for 40% and 10% base rate); AUC = ROC area under the curve, R^2 = Nagelkerke R^2; BIC = Bayesian information criterion; TOMM1 = Test of Memory Malingering–Trial 1; TOMM2 = Test of Memory Malingering–Trial 2; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.

(Table continues…)

	Hit Rate	Sn	Sp	PPP BR 40%	NPP BR 40%	PPP BR 10%	NPP BR 10%	R^2	AUC	AUC 95% CI	BIC
Two-Variable Models:											
TOMM2-C + TOMM1 RT correct	.92	.87	.95	.93	.91	.64	.99	.87	.98	[.97, 1.0]	-447.31
TOMM2-C + TOMM1 RT incorrect	.89	.82	.95	.91	.89	.64	.97	.75	.94	[.90, .99]	-336.86
TOMM2-C + TOMM2 RT correct	.94	.91	.97	.95	.94	.77	.99	.89	.99	[.98, 1.0]	-449.67
TOMM2-C + TOMM2 RT incorrect	.92	.92	.92	.90	.93	.56	1.0	.87	.99	[.98, 1.0]	-176.12
TOMM2-C + TOMM1 RT mean	.93	.89	.97	.95	.93	.75	.99	.88	.99	[.97, 1.0]	-447.65
TOMM2-C + TOMM2 RT mean	.93	.89	.97	.95	.93	.75	.99	.88	.99	[.97, 1.0]	-447.78
TOMM2-C + TOMM1 RT CV	.91	.80	.98	.97	.89	.80	.98	.75	.92	[.86, .98]	-423.17
TOMM2-C + TOMM2 RT CV	.91	.80	.98	.97	.89	.80	.98	.75	.89	[.81, .97]	-422.88
TOMM2-C + TOMM1 RT C/I	.89	.80	.98	.97	.88	.78	.98	.73	.91	[.85, .98]	-332.84
TOMM2-C + TOMM2 RT C/I	.94	.95	.92	.90	.97	.56	1.0	.89	.99	[.98, 1.0]	-178.05

Note. Sn = Sensitivity (detection of simulated TBI), Sp = Specificity (bona fide TBI); PPP = Positive Predictive Power, NPP = Negative Predictive Power (each presented for 40% and 10% base rate); AUC = ROC area under the curve, R^2 = Nagelkerke R^2; BIC = Bayesian information criterion; TOMM1 = Test of Memory Malingering–Trial 1; TOMM2 = Test of Memory Malingering–Trial 2; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.

Table 3b. *Logistic Regressions: TOMM Predicting HC and SIM Group Membership*.

	Df	X^2	*p*	*Odds Ratio*[3]	*Predictor*[1] *p*
One-Variable Models:					
TOMM1 Correct	1	94.29	< .001	0.63	
TOMM2 Correct	1	87.06	< .001	0.42	
TOMM1 RT correct	1	49.33	< .001	1.00	
TOMM1 RT incorrect	1	1.22	.270	1.00	
TOMM2 RT correct	1	63.86	< .001	1.01	
TOMM2 RT incorrect	1	0.22	.639	1.00	
TOMM1 RT mean	1	49.09	< .001	1.00	
TOMM2 RT mean	1	65.48	< .001	1.01	
TOMM1 RT CV	1	1.81	.179	4.55	
TOMM2 RT CV	1	12.57	< .001	307.66	
TOMM1 RT C/I[1]	1	12.38	< .001	14.99	
TOMM2 RT C/I[1]	1	6.66	.010	45.25	
Two-Variable Models:					
TOMM2-C + TOMM1 RT correct	2	111.50	< .001	1.01	.001
TOMM2-C + TOMM1 RT incorrect	2	74.21	< .001	1.00	.048
TOMM2-C + TOMM2 RT correct	2	113.87	< .001	1.01	.001
TOMM2-C + TOMM2 RT incorrect	2	45.29	< .001	1.00	.489
TOMM2-C + TOMM1 RT mean	2	111.85	< .001	1.01	.001
TOMM2-C + TOMM2 RT mean	2	111.97	< .001	1.01	< .001
TOMM2-C + TOMM1 RT CV	2	87.37	< .001	3.16	.568
TOMM2-C + TOMM2 RT CV	2	87.07	< .001	0.74	.915
TOMM2-C + TOMM1 RT C/I	2	70.19	< .001	1.14	.904
TOMM2-C + TOMM2 RT C/I	2	47.22	< .001	65.77	.197

Note. TOMM2-C = TOMM2 Correct; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.
1. Refers to RT-variable predictors added on the second step of the two-variable model.

Table 4a. *Classification Statistics: TOMM Performance for Single and Two-variable Models Predicting Simulator (SIM) and Traumatic Brain Injury (TBI) Group Membership*.

	Hit Rate	Sn	Sp	PPP BR 40%	NPP BR 40%	PPP BR 10%	NPP BR 10%	R^2	AUC	AUC 95% CI	BIC
One-Variable Models:											
13. TOMM1 Correct	.77	.73	.80	.70	.81	.30	.97	.48	.85	[.78, .93]	-311.24
14. TOMM2 Correct	.87	.78	.96	.93	.87	.70	.98	.67	.88	[.80, .96]	-333.85
15. TOMM1 RT correct	.63	.53	.73	.56	.70	.19	.94	.14	.64	[.52, .76]	-281.42
16. TOMM1 RT incorrect	.52	.96	.03	.40	.67	.10	1.0	.00	.54	[.41, .66]	-247.56
17. TOMM2 RT correct	.63	.49	.78	.60	.70	.18	.93	.20	.68	[.57, .79]	-285.76
18. TOMM2 RT incorrect	.67	1.00	.00	.40	.00	.11	.00	.05	.60	[.44, .76]	-151.73
19. TOMM1 RT mean	.58	.49	.67	.50	.67	.13	.92	.12	.63	[.51, .75]	-279.86
20. TOMM2 RT mean	.66	.53	.78	.61	.71	.22	.94	.21	.69	[.58, .80]	-286.57
21. TOMM1 RT CV	.50	.40	.60	.39	.59	.11	.91	.02	.57	[.45, .69]	-272.51
22. TOMM2 RT CV	.59	.49	.69	.51	.67	.14	.92	.05	.62	[.50, .74]	-274.43
23. TOMM1 RT C/I	.62	.69	.54	.50	.73	.15	.93	.18	.70	[.59, .81]	-259.08
24. TOMM2 RT C/I	.67	1.00	.00	.40	.00	.11	.00	.00	.53	[.35, .72]	-149.81

Note. Sn = Sensitivity (detection of simulated TBI), Sp = Specificity (bona fide TBI); PPP = Positive Predictive Power, NPP = Negative Predictive Power (each presented for 40% and 10% base rate); AUC = ROC area under the curve, R^2 = Nagelkerke R^2; BIC = Bayesian information criterion; TOMM1 = Test of Memory Malingering–Trial 1; TOMM2 = Test of Memory Malingering–Trial 2; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.

(Table continues…)

	Hit Rate	Sn	Sp	PPP BR 40%	NPP BR 40%	PPP BR 10%	NPP BR 10%	R^2	AUC	AUC 95% CI	BIC
Two-Variable Models:											
TOMM2-C + TOMM1 RT correct	.90	.82	.98	.97	.90	.78	.98	.71	.92	[.86, .98]	-334.98
TOMM2-C + TOMM1 RT incorrect	.88	.80	.97	.93	.88	.78	.97	.68	.92	[.86, .98]	-302.44
TOMM2-C + TOMM2 RT correct	.86	.78	.93	.88	.86	.54	.97	.71	.91	[.84, .97]	-334.36
TOMM2-C + TOMM2 RT incorrect	.91	.89	.95	.91	.91	.63	.98	.80	.97	[.86, .97]	-194.40
TOMM2-C + TOMM1 RT mean	.90	.82	.98	.97	.90	.78	.98	.71	.97	[.92, 1.0]	-334.92
TOMM2-C + TOMM2 RT mean	.84	.78	.91	.85	.86	.50	.97	.71	.92	[.85, .97]	-334.47
TOMM2-C + TOMM1 RT CV	.87	.78	.96	.93	.87	.70	.98	.67	.92	[.86, .97]	-329.35
TOMM2-C + TOMM2 RT CV	.87	.78	.96	.93	.87	.70	.98	.67	.87	[.78, .95]	-329.79
TOMM2-C + TOMM1 RT C/I	.86	.78	.95	.90	.87	.64	.97	.65	.90	[.83, .97]	-298.81
TOMM2-C + TOMM2 RT C/I	.91	.92	.89	.88	.94	.56	.98	.80	.88	[.80, .96]	-194.20

Note. Sn = Sensitivity (detection of simulated TBI), Sp = Specificity (bona fide TBI); PPP = Positive Predictive Power, NPP = Negative Predictive Power (each presented for 40% and 10% base rate); AUC = ROC area under the curve, R^2 = Nagelkerke R^2; BIC = Bayesian information criterion; TOMM1 = Test of Memory Malingering–Trial 1; TOMM2 = Test of Memory Malingering–Trial 2; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.

Table 4b. *Logistic Regressions Statistics: TOMM Predicting TBI and Simulator Group Membership*.

	Df	X^2	*p*	*Odds Ratio*	*Predictors*[1] *p*
One-Variable Models:					
TOMM1 Correct	1	40.03	< .001	1.26	
TOMM2 Correct	1	62.63	< .001	1.61	
TOMM1 RT correct	1	10.21	.001	1.00	
TOMM1 RT incorrect	1	0.25	.614	1.00	
TOMM2 RT correct	1	14.55	< .001	1.00	
TOMM2 RT incorrect	1	1.92	.165	1.00	
TOMM1 RT mean	1	8.64	.003	1.00	
TOMM2 RT mean	1	15.35	< .001	1.00	
TOMM1 RT CV	1	1.30	.255	0.22	
TOMM2 RT CV	1	3.22	.073	0.03	
TOMM1 RT C/I	1	11.77	.001	0.02	
TOMM2 RT C/I	1	0.003	.957	0.94	
Two-Variable Models:					
TOMM2-C + TOMM1 RT correct	2	68.27	< .001	1.00	.027
TOMM2-C + TOMM1 RT incorrect	2	59.56	< .001	1.00	.061
TOMM2-C + TOMM2 RT correct	2	67.65	< .001	1.00	.042
TOMM2-C + TOMM2 RT incorrect	2	48.64	< .001	1.00	.439
TOMM2-C + TOMM1 RT mean	2	68.21	< .001	1.00	.027
TOMM2-C + TOMM2 RT mean	2	67.76	< .001	1.00	.044
TOMM2-C + TOMM1 RT CV	2	62.64	< .001	1.07	.977
TOMM2-C + TOMM2 RT CV	2	63.08	< .001	8.75	.517
TOMM2-C + TOMM1 RT C/I	2	56.03	< .001	1.55	.798
TOMM2-C + TOMM2 RT C/I	2	48.44	< .001	0.31	.513

Note. TOMM2-C = TOMM2 Correct; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.
1. Refers to RT-variable predictors added on the second step of the two-variable model.

Table 5a. *Classification Statistics: TOMM Performance for Single and Two-variable Models Predicting Full Effort Healthy Comparison (HC) and Traumatic Brain Injury Group (TBI) Group Membership*.

	Hit Rate	Sn	Sp	PPP BR 40%	NPP BR 40%	PPP BR 10%	NPP BR 10%	R^2	AUC	AUC 95% CI	BIC
One-Variable Models:											
25. TOMM1 Correct	.76	.58	.90	.81	.76	.38	.96	.32	.77	[.68, .87]	-369.32
26. TOMM2 Correct	.63	.42	.79	.58	.67	.17	.93	.09	.62	[.51, .73]	-348.18
27. TOMM1 RT correct	.75	.67	.80	.68	.78	.27	.96	.39	.82	[.74, .90]	-377.13
28. TOMM1 RT incorrect	.49	.13	.82	.31	.59	.07	.90	.01	.57	[.44, .69]	-243.58
29. TOMM2 RT correct	.76	.64	.85	.73	.78	.33	.95	.41	.82	[.74, .90]	-379.38
30. TOMM2 RT incorrect	.66	1.00	.15	.45	1.0	.11	1.0	.02	.46	[.24, .68]	-61.11
31. TOMM1 RT mean	.74	.67	.79	.68	.78	.26	.96	.39	.81	[.73, .89]	-376.54
32. TOMM2 RT mean	.77	.64	.87	.77	.79	.37	.95	.40	.82	[.74, .90]	-378.25
33. TOMM1 RT CV	.58	.00	1.0	.00	.60	.00	.90	.00	.54	[.43, .65]	-340.53
34. TOMM2 RT CV	.61	.20	.88	.50	.62	.15	.91	.05	.69	[.59, .79]	-344.67
35. TOMM1 RT C/I	.51	.21	.77	.39	.59	.11	.89	.02	.62	[.50, .74]	-244.16
36. TOMM2 RT C/I	.66	.79	.46	.50	.75	.16	.93	.09	.63	[.43, .83]	20.06

Note. Sn = Sensitivity (detection of simulated TBI), Sp = Specificity (bona fide TBI); PPP = Positive Predictive Power, NPP = Negative Predictive Power (each presented for 40% and 10% base rate); AUC = ROC area under the curve, R^2 = Nagelkerke R^2; BIC = Bayesian information criterion; TOMM1 = Test of Memory Malingering–Trial 1; TOMM2 = Test of Memory Malingering–Trial 2; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.

(Table continues…)

	Hit Rate	Sn	Sp	PPP BR 40%	NPP BR 40%	PPP BR 10%	NPP BR 10%	R^2	AUC	AUC 95% CI	BIC
Two-Variable Models:											
TOMM2-C + TOMM1 RT correct	.77	.69	.84	.74	.80	.32	.96	.40	.82	[.74, .90]	-373.17
TOMM2-C + TOMM1 RT incorrect	.64	.46	.80	.60	.69	.21	.94	.10	.65	[.53, .77]	-244.98
TOMM2-C + TOMM2 RT correct	.77	.64	.87	.77	.79	.37	.95	.41	.82	[.74, .91]	-376.51
TOMM2-C + TOMM2 RT incorrect	.69	.89	.38	.48	.88	.14	1.0	.14	.68	[.49, .88]	-60.77
TOMM2-C + TOMM1 RT mean	.75	.62	.84	.72	.77	.32	.95	.40	.82	[.74, .90]	-372.82
TOMM2-C + TOMM2 RT mean	.77	.64	.87	.77	.79	.37	.95	.40	.82	[.74, .90]	-373.61
TOMM2-C + TOMM1 RT CV	.65	.36	.87	.65	.67	.25	.92	.10	.62	[.51, .73]	-343.58
TOMM2-C + TOMM2 RT CV	.64	.40	.82	.61	.68	.19	.93	.12	.70	[.61, .80]	-345.98
TOMM2-C + TOMM1 RT C/I	.63	.49	.75	.57	.69	.17	.93	.09	.67	[.56, .79]	-244.57
TOMM2-C + TOMM2 RT C/I	.59	1.0	.00	.41	.00	.09	.00	.25	.74	[.57, .92]	-63.74

Note. Sn = Sensitivity (detection of simulated TBI), Sp = Specificity (bona fide TBI); PPP = Positive Predictive Power, NPP = Negative Predictive Power (each presented for 40% and 10% base rate); AUC = ROC area under the curve, R^2 = Nagelkerke R^2; BIC = Bayesian information criterion; TOMM1 = Test of Memory Malingering–Trial 1; TOMM2 = Test of Memory Malingering–Trial 2; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.

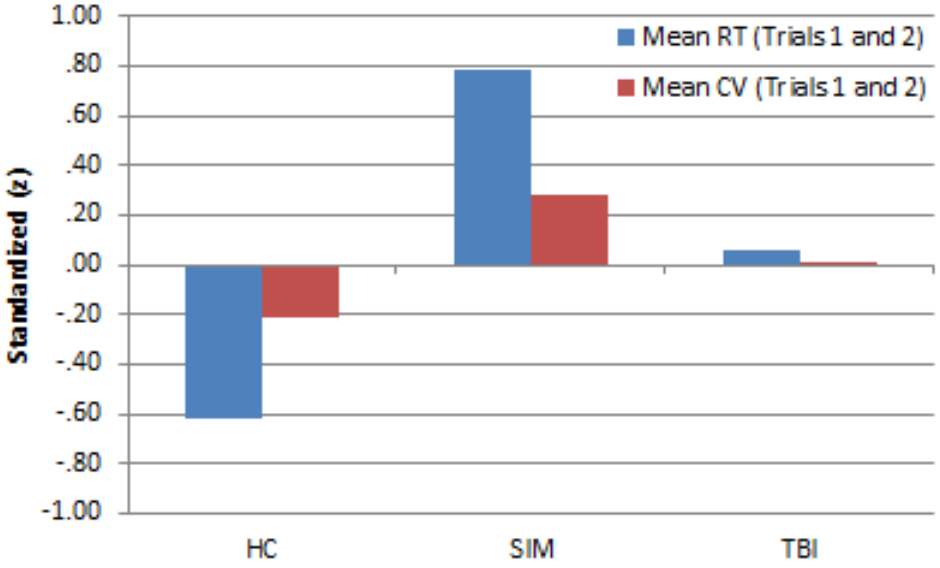
Table 5b. *Logistic Regressions: TOMM Performance Predicting HC and TBI Group Membership*.

	Df	X^2	*p*	*Odds Ratio*[3]	*Predictor*[1] *p*
One-Variable Models:					
TOMM1 Correct	1	28.84	< .001	0.75	
TOMM2 Correct	1	7.71	.006	0.53	
TOMM1 RT correct	1	36.66	< .001	1.00	
TOMM1 RT incorrect	1	0.41	.520	1.00	
TOMM2 RT correct	1	38.91	< .001	1.01	
TOMM2 RT incorrect	1	0.37	.541	1.00	
TOMM1 RT mean	1	36.07	< .001	1.00	
TOMM2 RT mean	1	37.77	< .001	1.00	
TOMM1 RT CV	1	0.06	.811	1.31	
TOMM2 RT CV	1	4.19	.041	27.33	
TOMM1 RT C/I[1]	1	0.99	.319	2.14	
TOMM2 RT C/I[1]	1	2.32	.128	5.39	
Two-Variable Models:					
TOMM2-C + TOMM1 RT correct	2	37.36	< .001	1.00	< .001
TOMM2-C + TOMM1 RT incorrect	2	6.23	.044	1.00	.422
TOMM2-C + TOMM2 RT correct	2	39.02	< .001	1.00	< .001
TOMM2-C + TOMM2 RT incorrect	2	3.49	.175	1.00	.304
TOMM2-C + TOMM1 RT mean	2	37.02	< .001	1.00	< .001
TOMM2-C + TOMM2 RT mean	2	37.8	< .001	1.00	< .001
TOMM2-C + TOMM1 RT CV	2	7.77	.021	1.35	.796
TOMM2-C + TOMM2 RT CV	2	10.17	.006	13.35	.123
TOMM2-C + TOMM1 RT C/I	2	5.82	.054	1.47	.627
TOMM2-C + TOMM2 RT C/I	2	6.46	.039	10.92	.065

Note. TOMM2-C = TOMM2 Correct; RT CV = Response Time coefficient of variation; RT C/I = Ratio RT correct / incorrect.
1. Refers to RT-variable predictors added on the second step of the two-variable model.

APPENDIX B (Figures)

Figure 1. TOMM Trials 1 and 2: Mean Response Time (RT) and Variability of Response Time (CV) for HC, SIM and TBI Groups



Group x Index interaction, $F(1, 148) = 11.44$, $p < .001$, $\eta^2 = .13$.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition). Washington D.C.: American Psychiatric Publishing.

Ashendorf, L., Constantinou, M., & McCaffrey, R.J. (2004). The effect of depression and anxiety on the TOMM in community-dwelling older adults. *Archives of Clinical Neuropsychology, 19*, 125–30.

Axelrod, B. N., Fichtenberg, N. L., Liethen, P. C., Czarnota, M. A., & Stucky, K. (2001). Performance characteristics of postacute traumatic brain injury patients on the WAIS-III and WMS-III. *The Clinical Neuropsychologist, 15*, 516–520.

Axelrod, B. N., Fichtenberg, N. L., Liethen, P. C., Czarnota, M. A., & Stucky, K. (2002). Index, summary, and subtest discrepancy scores on the WAIS-III in postacute traumatic brain injury patients. *International Journal of Neuroscience, 112*, 1479–1487.

Bashem, J.R., Rapport, L.J., Miller, J.B., Hanks, R.A., Axelrod, B.N., & Millis, S.R. (2014). Comparison of Five Performance Validity Indices in Bona Fide and Simulated Traumatic Brain Injury. *The Clinical Neuropsychologist, 28(5),* 851-875.

Bashem APA Bashem, J.R., Rapport, L.J., Billings, N.M., Miller, J.A., Hanks, R.A., Axelrod, B.N., & Kanser, R. (2014, August). Strategies of successful and unsuccessful simulators coached to feign traumatic brain injury. Presented at the 122st Annual Meeting of the American Psychological Association, Washington DC.

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,* 121-126.

Bender, S. D., & Rogers, R. (2004). Detection of neurocognitive feigning: Development of a

multi-strategy assessment. *Archives of Clinical Neuropsychology*, *19*, 49–60.

Bianchini, K.J., Mathias, C.W., & Greve, K.W. (2001) Symptom Validity Testing: A Critical Review, *The Clinical Neuropsychologist*, *15 (1),* 19-45

Binder, L. M., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *American Journal of Psychiatry, 153*(1), 7-10.

Bolan, B., Foster, J. K., Schmand, B., & Bolan, S. (2002). A comparison of three tests to detect feigned amnesia: The effects of feedback and the measurement of response latency. *Journal of Clinical and Experimental Neuropsychology*, *24*, 154–167.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R. & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN policy and planning committee. *Archives of Clinical Neuropsychology, 20(4),* 419–426.

Cicerone, K. D., Langenbahn, D. M., Braden, C., Malec, J. F., Kalmar, K., Fraas, M., ... Ashman, T. (2011). Evidence-based cognitive rehabilitation: Updated review of the literature from 2003 through 2008. Archives of Physical Medicine & Rehabilitation, 92, 519–530.

Coleman, R. D., Rapport, L. J., Millis, S. R., Ricker, J. H., & Falchion, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 20(2),* 201-210.

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluation children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology*, *9*(2), 81–89.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J.M., & McCaffrey, R.J. (2005). Is poor

performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology, 20(2),* 191-198.

Coronado, V.G., McGuire, L.C., Sarmiento, K., et al. (2012). Trends in Traumatic Brain Injury in the US and the public health response: 1995-2009. *Journal of Safety Research.* 43(4):299-307.

DenBoer, J. W., & Hall, S. (2007). Neuropsychological test performance of successful brain injury simulators. *The Clinical Neuropsychologist, 21(6),* 943-955.

Dikmen, S., Corrigan, J. D., Levin, H. S., Machamer, J., Stiers, W., & Weisskopf, M. G. (2009). Cognitive outcome following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 24*, 430–438.

Draper, K., & Ponsford, J. (2008). Cognitive functioning ten years following traumatic brain injury and rehabilitation. *Neuropsychology, 22*(5), 618-625.

Ekman, P., O'Sullivan, M., & Frank, M. (1999). A few can catch a liar. *Psychological Science*, 10, 263–266.

Essig, S. M., Mittenberg, W., Petersen, R. S., Strauman, S., & Cooper, J. T. (2001). Practices in forensic neuropsychology: Perspectives of neuropsychologists and trial attorneys. *Archives of Clinical Neuropsychology, 16*(3), 271-291.

Faul, M., Xu, L., Wald, M. M., & Coronado, V. G. (2010). Traumatic brain injury in the United States: Emergency Department visits, hospitalizations, and deaths 2002 - 2006. Atlanta, GA: Centers for Disease Control and Prevention: National Center for Injury Prevention and Control.

Faust, D. (1995). The detection of deception. Neurologic Clinics, 13, 255–265.

Finkelstein, E., Corso, P., Miller, T. (2006). *Incidence and Economic Burden of Injuries in the United States.* New York, NY: Oxford University Press.

Gouvier, W. D., Prestholdt, Ph.H., & Warner, M. S. (1988). A survey of common misconceptions about head injury and recovery. Archives of Clinical Neuropsychology, 3, 331–343.

Green, R., Melo, B., Christensen, B., Ngo, L. A., Monette, G., & Bradbury, C. (2008). Measuring premorbid IQ in traumatic brain injury: An examination of the validity of the Wechsler Test of Adult Reading (WTAR). *Journal of Clinical and Experimental Neuropsychology, 30(2),* 163-172.

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

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

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. *Clin Neuropsychol.* 2008;22(5):896-918.

Greiffenstein, M.F., Gola, T., & Baker, W.J. (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. *The Clinical Neuropsychologist, 9,* 230-240

Gunstad, J., Suhr, J., (2001). Efficacy of the full and abbreviated forms of the Portland Digit Recognition Test: Vulnerability to coaching. *The Clinical Neuropsychologist, 15*, 397–404.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (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.

Hosmer, D. W., & Lemeshow, S. (2000). Applied logisitic regression. New York, NY: Wiley Interscience.

Iverson, G. L. (1995). Qualitative aspects of malingered memory deficits. *Brain Injury, 9*(1), 35-40.

Johnstone, B., Hexum, C., & Ashkanazi, G. (1995). Extent of cognitive decline in traumatic brain injury based on estimates of premorbid intelligence. *Brain Injury. 9(4),* 377-384.

Kaiman, R. (2003). Median award for mild brain injuries is $271,350. Personal Injury Verdict Reviews, 11, 3.

Langlois, J.A., Rutland-Brown, W., Wald, M.M. (2006) The epidemiology and Impact of Traumatic Brain Injury: A Brief Overview. *J Head Trauama Rehab. 21*(5) 375-378.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17(3),* 410-425.

Larrabee, G. J. (2005). Mild Traumatic Brain Injury. In G. J. Larrabee (Ed.), *Forensic Neuropsychology: A scientfic approach*. New York, NY: Oxford University Press.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *Clinical Neuropsychologist, 22*(4), 666-679.

Larrabee, G.J., Millis, S.R., & Meyers, J.E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *Clin Neuropsychol*; 23(5):841-849.

Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. Archives of Clinical Neuropsychology, 18, 261–276.

Meyers, J.E., Volbrecht, M., Axelrod, B.N., & Reinsch-Boothby, L. (2011). Embedded Symptom Validity Tests and Overall Neuropsychological Test Performance. *Archives of Clinical Neuropsychology.* 26(1):8-15.

Miele, A.S., Gunner, J.H., Lynch J.K., McCaffrey, R.J. (2012) Embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology,* 27, 10-22.

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

Miller, J.B., Millis, S.R., Rapport, L.J., Bashem, J.R., Hanks, R.A., & Axelrod, B.N. (2011). Detection of Insufficient Effort Using the Advanced Clinical Solutions for the Wechsler Adult Intelligence Scale, Fourth Edition. *Clinical Neuropyshcologist, 25*(1): 160-172.

Millis, S.R. (2008). Assessment of incomplete effort and malingering in the neuropsychological evaluation. In J. Morgan J. & J. Ricker (Eds.). *Textbook of Clinical Neuropsychology: Studies on Neuropsychology, Neurology and Cognition* (pp. 891-904). New York: Taylor & Francis.

Millis, S.R. (2009). What clinicians really need to know about symptom exaggeration, insufficient effort, and malingering: Statistical and measurement matters. In *Neuropsychology of malingering casebook*. New York: American Academy of Clinical

Neuropsychology.

Neudecker, J.J., Skeel, R.L. (2009). Development of a Novel Malingering Detection Method Involving Multiple Detection Strategies. *Archives of Clinical Neuropsychology. 24*, 59-70.

Orman J.L., Kraus, J. F., Zaloshnja, E., & Miller, T. (2011). Epidemiology. In: J. M. Silver TWM, & S. C. Yudofsky, ed. *Textbook of traumatic brain injury*. 2nd ed. Washington, DC: American Psychiatric Publishing, Inc.

Raftery, A. E. (1995). Bayesian model selection in social research. Sociological Methodology, 25, 111–163.

Rapport, L. J., Farchione, T. J., Coleman, R. D., & Axelrod, B. N. (1998). Effects of coaching on malingered motor function profiles. *Journal of Clinical Experimental Neuropsychology, 20(1),* 89-97.

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.

Rees, L.M., Tombaugh, T.N., & Boulay, L. (2001). *Depression and the Test of Memory Malingering. Archives of Clinical Neuropsychology, 16*, 501–6.

Reicker, L.I. (2008). The ability of reaction time tests to detect simulation: An investigation of contextual effects and criterion scores. *Archives of Clinical Neuropsychology, 23,* 419-431.

Rogers, R. (1990). Models of feigned mental illness. *Professional Psychology: Research and Practice, 21*, 182-188.

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

Rose, F. E., Hall, S., & Szalda-Petree, A. D. (1995). Portland Digit Recognition Test–

Computerized: Measuring response latency improves the detection of malingering. *The Clinical Neuropsychologist, 9*, 124–134.

Ruff, R. M., & Richardson, A. M. (1999). Mild Traumatic Brain Injury. In J. J. Sweet (Ed.), *Forensic Neuropsychology: Fundamentals and practice*. Lisse, Netherlands: Swets and Zeitlinger.

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(5),* 693-702.

Sharland, M. J. & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology 22(2),* 213-223.

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

Slick, D. J., Tan, J. E., et al. (2004). Detecting malingering: a survey of experts' practices. *Archives of Clinical Neuropsychology 19(4),* 465-473.

Suhr, J. A., & Barrash, J. (2007). Performance on Standard Attention, Memory, and Psychomotor Speed Tasks as Indicators of Malingering. In G. J. Larrabee (Ed.), *Assessment of Malingered Neurocognitive Deficits* (pp. 131 - 170). New York: Oxford University Press.

Suhr, J. A., & Gunstad, J. (2007). Coaching and Malingering: A Review. In G. J. Larrabee (Ed.), *Assessment of Malingered Neurocognitive Deficits* (pp. 287 - 311). New York: Oxford University Press.

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

Tabachnick, B.G. & Fidell, L.S. (2012). *Using Multivariate Statistics* (6th ed.). Boston, MA: Pearson Education, Inc.

Tombaugh, T. (1996). *Test of Memory Malingering.* (1996). North Tonawanda, NY: Multi-Health Systems.

Tombaugh, T.N. (1997). The test of memory malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment.* 9(3):260-268.

Vagnini, V.L., Berry, D.T.R., Clark, J.A., & Jiang, Y. (2008). New measures to detect malingered neurocognitive deficit: Applying reaction time and event-related potentials. *Journal of Clinical and Experimental Neuropsychology, 30,* 766-776

van Hooff, J.C., Sargeant, E., Foster, J.K., & Schmand, B.A. (2009). Identifying deliberate attempts to fake memory impairment through the combined use of reaction time and event-related potential measures. *International Journal of Psychophysiology, 73,* 246-256.

Willison, J., & Tombaugh, T. N. (2006). Detecting simulation of attention deficits using reaction time tests. *Archives of Clinical Neuropsychology*, *21*, 41–52.

Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J., & Sweet, J. J. (2010). Effort Indicators within the California Verbal Learning Test-II (CVLT-II). Clinical Neuropsychologist, 24(1), 153-168

ABSTRACT

DETECTION OF MALINGERING IN BONA FIDE TRAUMATIC BRAIN INJURY AND SIMULATED TRAUMATIC BRAIN INJURY: COMBINING RESPONSE TIME WITH PVT ACCURACY RESULTS

by

ROBERT KANSER

May 2016

Advisor: Dr. Lisa J. Rapport

Major: Psychology (Clinical)

Degree: Master of Arts

Threats to performance validity test (PVT) security and utility have increased efforts to develop covert measures of performance validity. Response time (RT) is a promising covert measure to distinguish between honest and feigned performance; however, research investigating RT patterns on PVTs is sparse and troubled by methodological problems. This study examined the incremental utility of RT variables on a computerized version of the Test of Memory Malingering (TOMM-C) in distinguishing adults with verified traumatic brain injury (TBI) and healthy adults coached to feign neurocognitive impairment. Participants were 45 adults with moderate to severe TBI, 45 healthy adults coached to feign neurocognitive impairment (SIM), and 61 healthy adult comparisons providing full effort (HC). A number of RT indices differed significantly across groups. RT indices and traditional TOMM-C accuracy scores were evaluated using logistic regression, ROC curve, and Bayesian Information Criterion statistics. Mean RT on Trial 1 and 2 provided incremental predictive value to traditional TOMM-C accuracy in discriminating groups (SIM vs. HC and SIM vs. TBI). Degree of preference for RT indices depended on which groups were being discriminated.

AUTOBIOGRAPHICAL STATEMENT

ROBERT KANSER

Education

08/2008 - 05/2012 **Bachelor of Science**
University of Michigan
Major: Biopsychology, Cognition, and Neuroscience

Clinical Experience

August 2013 – Present **Wayne State University Psychology Clinic, Detroit, MI**
Individual psychological assessment and psychotherapy
Interpersonal group therapy co-leader

August 2015 **Confident Kids Camp, Ann Arbor, MI**
Thirty hours of intensive individual behavioral therapy for children with Selective Mutism
Advisor: Aimee Kotrba, Ph.D.

September 2015 – Present **Center for Forensic Psychiatry, Ann Arbor, MI**
Psychological assessment and group psychotherapy for inpatients found incompetent to stand trial or not guilty by reason of insanity
Dialectical Behavior Therapy group co-leader
Advisors: Judith Shazer, Ph.D., and Jay Witherell, Ph.D.

Research Experience

May 2012 – June 2013 **Research Assistant**
Traumatic Brain Injury Clinic and Community Living Center (inpatient)
Department of Veteran Affairs-Medical Center, Ann Arbor, MI
Advisor: Linas Bieliauskas, Ph.D.

January 2012 – May 2015 **Research Assistant**
Affective Neuroscience & Biopsychology Lab
University of Michigan Psychology Department, Ann Arbor, MI
Advisor: Kent Berridge, Ph.D.

August 13 – Present **Research Assistant**
Wayne State University Psychology Department, Detroit, MI
Advisor: Lisa Rapport, Ph.D.

Honors and Awards

2010 - 2012 James B. Angel Scholar
University of Michigan, Ann Arbor, MI

08/2012 Phi Beta Kappa Honor Society
University of Michigan, Ann Arbor, MI