January 2019

The Eyes Never Lie: Detecting Simulated Traumatic Brain Injury With Eye-Tracking

Robert John Kanser
Wayne State University, robert.kanser@wayne.edu

THE EYES NEVER LIE:
DETECTING SIMULATED TRAUMATIC BRAIN INJURY WITH EYE-TRACKING

by

ROBERT J KANSER

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2019

MAJOR: CLINICAL PSYCHOLOGY

Approved By:

_____

Advisor Date

_____

_____

_____

_____

ACKNOWLEDGMENTS

This project was far from a solitary endeavor. It is the result of years of training, mentorship, and collaboration. I owe special thanks to a number of people who made completion of this project possible. I would like to express my deepest gratitude to my advisor, Dr. Lisa Rapport, for her constant encouragement, guidance, and support throughout this project and my professional development as a whole. I could not have asked for a better mentor. I have been incredibly fortunate to train and work under a number of excellent mentors and role models, including members of my dissertation committee. Thank you to Drs Robin Hanks, Linas Bieliauskas, and Douglas Whitman for contributing your time, expertise, and consultation to this project. This project would not have been possible without the efforts and collaboration of our research team: Drs Jesse Bashem and Rachel Keelan, Nia Billings, Monica De Iorio, Sarah Patrick, Barret Vermillion, Rob Kotasek, and Carole Koviak. Lastly, I must express my deepest thanks to my friends and family. Their constant love and support is a driving force in all I do.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

As public knowledge of TBI and its associated deficits grows, there comes an increased risk of individuals attempting to feign neurocognitive impairment for secondary gain such as financial incentives or mitigating responsibility. Estimated base rates of feigned neurocognitive impairments are 30-50% for litigation or compensation seeking cases (G. J. Larrabee, Millis, & Meyers, 2009a). Clearly, the ability to distinguish actual TBI from feigned neurocognitive impairment is critical to both the healthcare and legal systems. A great deal of research has been directed at developing *performance validity tests* (PVTs) designed to identify feigned or exaggerated presentations of impairment during neurocognitive evaluations. An unfortunate similarity among PVTs is that most display only moderate sensitivity to detect individuals who purposefully perform poorly (Bianchini, Mathias, & Greve, 2001; Lippa, 2018). Improving the diagnostic sensitivity of PVTs without sacrificing specificity that protects patients from false diagnoses of malingering is an essential but complex goal. Among the most central constraints on this endeavor are the research paradigms traditionally used to investigate PVTs. This study focused on test and extra-test (research design) characteristics that can enhance the validity of PVTs.

Section 1. Assessment of Traumatic Brain Injury in Medicolegal Settings

Traumatic brain injury (TBI) is a prominent and prevalent health concern that is associated with a variety of behavioral and cognitive deficits. The severity and duration of these deficits vary greatly and depend on a number of factors. Neuroimaging plays an important role in identifying brain injury; however, neuroimaging may not reveal structural abnormalities in individuals with verified TBI. Furthermore, neuroimaging is not, nor was it ever, intended to predict cognitive and functional deficits (Page, 2006; Wasyliw & Golden, 1985).

Neuropsychological assessments play a critical role in diagnosing TBI, describing and delineating cognitive and functional deficits, and monitoring rehabilitation. In doing so, neuropsychological assessments help ensure that individuals with TBI receive the most useful rehabilitative treatments, and that resources are allocated to persons who truly need them.

The validity and utility of information obtained through neuropsychological assessments 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 (Green, Rohling, Lees-Haley, & Allen, 2001; Meyers, Reinsch-Boothby, Miller, Rohling, & Axelrod, 2011). An examinee providing suboptimal effort could, therefore, receive test scores that indicate substantially greater deficits than they truly have. Consequently, examinees may receive unnecessary, ineffective, and/or undue treatment services and resources. Unfortunately, feigned impairment or purposeful suboptimal effort associated with TBI is common, especially in compensation and litigation settings. As such, research focused on creating standardized tests to assess validity of effort during neurocognitive evaluation (performance validity) has burgeoned.

The most common PVT paradigm that dominates neuropsychological assessments today was developed over 40 years ago (Bianchini et al., 2001; Pankratz, Fausti, & Peed, 1975). Although advancements have been made, PVTs have shown limited evolution. Moreover, PVT development has been insufficiently informed by neurocognitive processing theories and uncoupled with advancements in neuroscience and modern technologies (Leighton, Weinborn, & Maybery, 2014). This lack of evolution has resulted in a lack of PVT diversity. A large number of PVTs exist today; however, the vast majority are strikingly similar, sharing a common structure, target construct, and scoring strategy.

Overall, classification accuracy of PVTs in distinguishing actual TBI from feigned impairment remains error prone. Research has investigated test and extra-test (e.g., research design) characteristics that likely contribute to PVTs inaccuracies. The extra-test variables have received limited empirical study, and findings from those studies have been underutilized. In contrast, test characteristics of PVTs that contribute to their inaccuracies have been relatively well studied. These limitations are compounded by the fact that information that threatens PVT test security is readily available on the internet and from informed attorneys (Bauer & McCaffrey, 2006). Clearly, there is a need for adaptation and modernization within the field of performance validity assessment.

The proposed study seeks to utilize a unique, ecologically valid research design to address limitations common to most PVT research. Doing so will facilitate achievement of the primary aim of the proposed study, to determine the incremental utility of combining covert measures, informed by neurocognitive processing theory, with established PVTs. Specifically, this study will examine the extent to which analysis of oculomotor patterns on a computerized version of the Warrington Recognition Memory Test of Words (Warrington, 1984) may improve its classification accuracy in distinguishing TBI simulators from individuals with verified 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. 76). 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, but the incentive is thought to be internal/psychological (i.e., to play the sick role and/or receive attention). Symptom presentation in *conversion disorder* is thought to be unconscious, and incentive also is considered internal/psychological (i.e. to 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 (i.e., 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, 1994; Rogers, 1990).

Slick et al. (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. (1999) 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 multimethod, multidimensional approach that requires the integration of data from self-reported symptoms, medical histories, behavioral observations, and neuropsychological testing.

Slick et al. (1999) note that, even with explicit, reliable criteria that integrate multiple 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. *Definite MND* requires evidence of negative response bias, or worse than chance performance on a forced-choice task. *Probable* malingering must include either two or more types of evidence from neuropsychological testing or one piece of evidence from neuropsychological testing and one or more pieces of evidence from self-report. *Possible* malingering is used when there is only evidence from self-report or instances where psychiatric, neurologic, or developmental factors cannot be ruled out. Neuropsychological test evidence includes failure of a well-validated PVT and discrepancies between test data and known patterns of brain functioning, observed behavior, reliable collateral reports, or documented history/background. Similarly, self-report evidence

includes identification by a well-validated symptom validity scale/measure and discrepancies between self-report and the previously mentioned sources of information.

Slick and Sherman (2012) proposed a number of revisions to the Slick et al. (1999) criteria in an attempt to address concerns and limitations associated with MND. As the term MND suggests, emphasis was placed on the malingering of neurocognitive symptoms; however, malingering can extend to a variety of neurobehavioral and neuropsychiatric symptomology. Accordingly, the term malingered neuropsychological dysfunction was proposed as a more comprehensive and accurate alternative. Additionally, the previous criteria valued evidence of exaggeration from objective test data over that obtained through self-report data; however, research does not support this distinction (G. J. Larrabee, Greiffenstein, M.F., Greve, K.W. & Bianchini, K.J., 2007). For example, diagnosis of *definite* MND could only be made following below chance performance on a forced-choice PVT. Evidence sufficient for *definite* malingered neuropsychological dysfunction now includes: below chance performance on a forced-choice PVT, high probability (> .95) of exaggeration based on the single or combined predictive probability of one or more reliable sources (e.g., PVT failures, SVT failures, etc.), and self-reported symptoms that are "unambiguously incompatible with or directly contradicted by directed observed behavior and/or test performance."

Next, the exclusion criteria from the Slick et al. (1999) criteria were removed given concern that they implied individuals with severe psychiatric, neurological, or developmental disorders are incapable of malingering. Secondary malingered neuropsychological deficit now describes individuals who meet criteria for definite or probable malingering, but presentation is attributable to genuine deficits that prevent the individual from controlling their behavior or understanding its moral implications (e.g., an individual with schizophrenia exaggerating as a

result of delusions or command hallucinations; Slick & Sherman, 2012). Finally, the new criteria also distinguish the previously explored cases of malingering by proxy (Cassar, Hales, Longhurst, & Weiss, 1996; Lu & Boone, 2002). Malingered neuropsychological dysfunction by proxy occurs when presentation is due to significant influence, control, and/or instruction from an external party (e.g., a parent directing a child's presentation, an abusive partner threatening violence if symptoms are not exaggerated, etc.).

Section 1.2 – Clinical Significance

Traumatic brain injury is the leading cause of disability in individuals under the age of 40 (Draper & Ponsford, 2008). Although TBI contributes to approximately 30% of all injury-related deaths in the United States each year, medical advancements have resulted in a larger number of individuals surviving TBIs that would have been lethal in the past (Faul, Xu, Wald, & Coronado, 2010). There is an estimated 5.3 million people living with TBI related deficits in the United States alone (Langlois, Rutland-Brown, & Wald, 2006). In 2013, there were an estimated 2.8 million TBI-related emergency room (ER) visits, hospitalizations, and deaths (Taylor, Bell, Breiding, & Xu, 2017). Since 1991, these TBI-related incidents have increased (Coronado et al., 2012; Taylor et al., 2017). The 50% increase in TBI-related ER visits was largely due to increased falls in adults older than the age of 75 (Taylor et al., 2017). The statistical rise is also likely due, in part, to 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 (G. J. 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). Clearly, the risk and prevalence of malingering in the context of forensic traumatic brain injury assessment is great.

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 (G. J. Larrabee, Millis, & Meyers, 2009b). With such high prevalence, it is clear that the ability to assess malingering accurately is critical, especially in areas where external incentives are common.

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 (Corso, Finkelstein, Miller, Fiebelkorn, & Zaloshnja, 2006; Langlois Orman, Kraus, Zaloshnja, & Miller, 2011). Accounting for indirect costs of TBI, such as loss of productivity, estimates exceed $76.5 billion annually (Corso et al., 2006; Corso, Finkelstein, Miller, Fiebelkorn, & Zaloshnja, 2015). 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 distinguish between

verified TBI and feigned cognitive impairment has substantial economic and social consequences for patients, the healthcare system, and the legal system.

Section 1.3 – Evolution of Effort Assessment

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 variety of sources. As highlighted by the Slick criteria, one fundamental approach is to examine discrepancies between test or self-report data and reliable sources of information (e.g., known patterns of brain functioning, reliable collateral reports, etc.). This type of qualitative discrepancy analyses, however, relies heavily on clinical judgment. Unfortunately, research has consistently shown clinicians are poor predictors of invalid patient presentation/performance when using combinations of test data, information gained through clinical interview/medical record review, and behavioral observations (Dandachi-FitzGerald, Merckelbach, & Ponds, 2017; Ekman, O'Sullivan, & Frank, 1999; Faust, 1995; Heaton, Smith, Lehman, & Vogt, 1978). Findings such as these highlight the need for measures of effort that emphasize quantitative versus qualitative aspects.

Tests of effort have been called many things (e.g., malingering tests, tests of response bias, symptom validity tests, etc.). G. J. 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 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, Bauer, Ashendorf, Fisher, & McCaffrey, 2005; Millis, 2008). Accordingly, PVT usage in nearly all neuropsychological evaluations has been deemed a standard of practice by leading neuropsychological organizations (Chafetz et al., 2015; Heilbronner et al., 2009; Sweet, Benson, Nelson, & Moberg, 2015).

Modern performance validity assessment can trace its origin back over 40 years (Bianchini et al., 2001). Pankratz et al. (1975) first designed a forced-choice task to identify feigned deafness. In general, forced-choice tasks present a simple target stimulus (e.g., sounds, words, numbers, etc.) followed by a two-item forced-choice recognition task in which the target stimuli are paired with foils. Individuals must correctly identify the target stimuli. Through the use of a forced-choice format, Pankratz et al. (1975) were able to compare participant performance to what would be expected based on chance alone. Specifically, participants should be able to achieve approximately 50% accuracy if they were simply guessing. Performance significantly below chance, or a negative response bias, provides strong evidence that the individual is able to identify the correct response but is deliberately responding incorrectly. Pankratz is credited with naming this process 'symptom validity testing' (now performance validity testing) and extending the procedure beyond feigned perceptual deficits and into the assessment of feigned memory impairment (Bianchini et al., 2001; Pankratz, 1983). Pankratz developed subject-specific forced-choice tasks based on the subjects' self-reported memory difficulty (i.e., inability to remember names, addresses, etc.). Unfortunately, these tasks inherently lacked standardization due to their high degree of specificity.

A number of different standardized PVTs have been created since 1975; however, PVTs have shown limited evolution. Modern PVTs are strikingly similar and share a common set of features. 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, Prestholdt, & Warner, 1988). Furthermore, feigned memory impairment is among the most common strategies used by individuals instructed to simulate TBI (Iverson, 1995; Kanser et al., 2017; Tan, Slick, Strauss, & Hultsch, 2002). Additionally, the most commonly used and best-validated PVTs employ a similar structure; a two-item forced-choice paradigm that enables detection of negative response bias. Minor structural differences include type (e.g., numbers vs. words vs. line drawings) and number of stimuli, number of stimuli presentations and forced-choice trials, and potential delay trials (Bianchini et al., 2001).

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 adults and individuals with verified 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 (Backhaus, Fichtenberg, & Hanks, 2004; Bender & Rogers, 2004; Neudecker & Skeel, 2009). Doing so increases sensitivity (i.e., the proportion of individuals providing suboptimal effort correctly identified as such by the test) while maintaining a clinically acceptable specificity (i.e., the proportion of individuals providing adequate effort correctly identified as such). Unfortunately, the most commonly used and best-validated PVTs show only moderate sensitivity (Bianchini et al., 2001; Sollman & Berry, 2011).

Research has investigated the test and extra-test (e.g., research design components) variables that may account for PVTs inaccuracies.

Section 1.4 – Research Design and Limitations

A major factor contributing to the moderate sensitivity of most PVTs are the limitations of current research designs used in PVT development, evaluation, and evolution. Several basic research designs exist, each with their own advantages and limitations that influence how commonly they are employed. *Case studies* provide rich descriptive information that facilitate clinical case conceptualization, which can play an important role in the early phases of PVT development. Case studies are rarely used in clinical research, however, due to significant concerns over generalizability (G. J. Larrabee, 2011). *Differential prevalence designs*, another infrequently used method, compare groups with differential base rates of malingering (i.e., litigating and non-litigating patients). Interpretation of results from differential prevalence designs is tenuous, as group classification is not based on empirical evidence and, therefore, rates of feigned impairment in groups are unknown (Rogers, 1997). Clearly, not all individuals in litigation are feigning impairment and, based on previous literature, it is safe to conclude that a portion of non-litigating patients is feigning impairment.

Known-group designs use objective criteria to define clinical groups of interest for comparison. In the context of PVT research, these groups typically include clinical patients who meet criteria for malingering and patients with verified brain injury who do not meet criteria for malingering. The primary strength of this type of design is its generalizability, as it utilizes genuine clinical populations. A major limitation of known-group designs is that random assignment is lost. Group classifications are made using objective criteria, typically after the data have been collected (G. J. Larrabee, 2011). Moreover, even with the use of objective criteria, like

Slick et al. (1999) criteria for probable MND, error variance due to group misclassifications still exist. In other words, a portion of patients in the malingering group will be false positives (i.e., true brain injury, falsely labeled malingerers), and a portion of individuals in the full-effort TBI group will be false-negatives (i.e., sophisticated malingerers who have avoided detection). It is possible that neuroimaging technologies will continue to evolve, such that more patients with actual brain injuries will be verified and fewer will be missed and mislabeled (i.e., lowering false positives). In contrast, error variance from successful malingerers who have gone undetected by PVTs (1 – sensitivity) is practically unfeasible to eliminate in clinical settings. It will persist unless all such examinees offer honest posttest confessions that they feigned brain injury during testing—which will not happen given current societal reward and punishment systems.

Analog design evolved to address the error variance introduced by successful malingerers who go undetected by PVTs. It is the most common research design in the study of malingering and performance validity (Bianchini et al., 2001). In most analog designs, healthy college students are randomly assigned to one of two groups: those instructed to perform to the best of their ability, and those instructed to feign TBI. The impracticable task of quantifying unknown error variance among suspected malingerers in the clinical setting is converted to a form of known-group design in an experimental setting. The major strength of the analog design is the high degree of experimental control: Unlike the clinical setting, analog design removes uncertainty regarding whether the respondent was feigning TBI. Accordingly, it is especially useful in the early stages of PVT development (Bianchini et al., 2001).

Limitations of the analog design have been discussed in detail (G. J. Larrabee, 2005; Rogers, 2008; Suhr & Gunstad, 2007), but generally focus on the design's questionable ecological validity. Characteristics found to differentiate healthy adults providing full effort and

TBI simulators will not necessarily generalize to groups of clinical interest (i.e., individuals with verified TBI and those feigning impairment). In efforts to increase generalizability of findings, analog studies have begun to incorporate a group with verified TBI. Additional concerns surround the extent to which individuals instructed to simulate brain injury accurately represent clinical malingerers, given key differences between these groups. First, clinical malingerers have a tremendous financial incentive to appear brain injured and, therefore, a much greater motivation to do so successfully. Next, clinical malingerers traditionally have more time than do analog research participants to prepare for assessments and utilize resources that threaten PVT security/utility (e.g., the internet, legal counsel). Surveys of practicing attorneys found that almost 50% believe they should provide specific information about tests in the neurocognitive evaluation (including PVTs) 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, Petersen, Strauman, & Cooper, 2001). Lastly, and rarely considered, most test batteries in analog studies lack external validity because they include one or more PVTs but no tests of neurocognitive functioning (Suhr & Gunstad, 2007). As such, simulators in many analog designs are asked to do the impossible: appear brain injured on a test battery that consists only of PVTs that do not assess neurocognitive abilities.

Successful simulation of brain injury occurs when an individual performs in such a way as to appear impaired on cognitive tests, while at the same time avoiding detection of performance invalidity by PVTs. Successful malingerers must titrate performance in a narrow window, passing PVTs but presenting with some form of cognitive impairment that would warrant compensation. Thus, it is very possible that a compensation-seeking examinee who is feigning TBI would avoid malingering detection by passing the PVTs but then not perform

poorly enough on the neuropsychological tests within the cognitive evaluation to appear brain injured. The concept of *successful simulation* has not been well represented in PVT research (Kanser et al., 2017). Clearly, there is a need to adapt the analog design so that PVTs can be evaluated and refined in a context that more closely mirrors the challenges presented in clinical and forensic settings.

Section 1.5 – Research Design Solutions

Despite the heavy reliance on analog designs in the development and evaluation of PVTs, design manipulations that may enhance generalizability have been understudied and underutilized. Methods to enhance simulation sophistication have included providing simulators information about common symptoms associated with TBI (Coleman, Rapport, Millis, Ricker, & Farchione, 1998), warning about the presence of PVTs (Suhr & Gunstad, 2000), and providing general coaching on strategies to avoid detection (i.e., miss more difficult items than easy ones, don't miss more than half the questions; Brennan et al., 2009; Russeler, Brett, Klaue, Sailer, & Munte, 2008; Weinborn, Woods, Nulsen, & Leighton, 2012). Each of these design manipulations have been shown to enhance sophistication in the simulation presentation. In fact, providing information about symptoms of TBI in conjunction with test-specific coaching strategies appears to result in the most sophisticated TBI simulation (Suhr & Gunstad, 2007; Weinborn et al., 2012).

Section 1.6 – PVT Limitations

Another key factor contributing to the moderate sensitivity observed in the most widely used PVTs are test variables, or characteristics of the PVTs themselves. First and foremost, PVTs use cutoff scores that maximize specificity at the expense of sensitivity (Bianchini et al., 2001). 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.

In addition to this self-imposed restriction, PVTs appear to have relatively high face validity. In post-exam interviews, Tan et al. (2002) found that the vast majority of college students instructed to simulate brain injury correctly identified three of the most commonly used PVTs as measures of performance validity: Over 70% identified the Word Memory Test (Green, Allen, & Astner, 1996), over 75% identified the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997), and over 90% identified the Test of Memory Malingering (TOMM; Tombaugh, 1996) as, in part, measuring performance validity. It is important to note that participants were given a week to prepare for testing, and the most common self-reported preparation strategy was searching the internet. Information that jeopardizes PVT security is readily available on the internet (Bauer & McCaffrey, 2006). Searching basic terms (e.g., malingering, neuropsychological/independent medical evaluations, faking on neuropsychological tests) can provide information that facilitates feigning impairment (Ruiz, Drake, Glass, Marcotte, & van Gorp, 2002). Bauer and McCaffrey (2006) simply searched the TOMM and VSVT by name on Google and found that over 20% of the first 50 sites were significant threats to test security (i.e., described test format, cited studies examining the test, explained scoring, or even provided specific cut scores). Concerns over face validity and internet-based threats to PVT test-security are more problematic when one considers the fact that PVTs are susceptible to even the most elementary forms of coaching (e.g., warning about the presence of PVTs, providing information about common symptoms of TBI; Suhr & Gunstad, 2007).

Lastly, performance validity assessment as a whole is limited by the fact that the majority of PVTs are related to aspects of memory performance. Because most PVTs target memory

dissimulation, they may not be sensitive to individuals targeting other cognitive domains (e.g., attention, processing speed, etc.) or specific types of tests (e.g., visual or verbal) to feign cognitive impairment. Unfortunately, research regarding the qualitative strategies individuals employ when instructed to simulate neurocognitive impairment is sparse. In addition to memory tasks, the most successful simulators (i.e., those who perform impaired on cognitive tests and avoid PVT detection) commonly target visual tasks as a whole for symptom dissimulation. Additionally, slowed responding and deliberate inattention/poor concentration are two of the most common simulation techniques (Kanser et al., 2017; Tan et al., 2002). As such, an increasing amount of research has been directed at incorporating new, covert measures of effort into established PVTs. Two covert measures of effort that are uniquely sensitive to these strategies and tasks are response time and eye-tracking data.

Section 1.7 – Response Time & Latency

Response time (RT) has been identified as a promising covert measure to distinguish between honest and feigned performance. Clearly, RT would be sensitive to individuals employing a strategy of slowed responding. In addition to detecting this conscious, deliberate strategy, RT may serve as a marker sensitive to the cognitive processes underlying malingering. From a cognitive processing perspective, individuals malingering on a forced-choice recognition test must not only identify the correct item, but also decide whether or not to respond correctly (Bolan, Foster, Schmand, & Bolan, 2002). This added cognitive demand may unconsciously prolong RT and increase RT variability. Not surprisingly, RT has been shown to be more resistant to coaching than performance accuracy (Rose, Hall, Szalda-Petree, & Bach, 1998).

Computer-based RT tasks have been developed with the goal of distinguishing feigned neurocognitive impairment from verified TBI. Generally, these tasks involve a simple RT task in

which participants must respond as quickly as possible when some stimuli (visual or auditory) is presented. Some RT tests, like the Computerized Tests of Information Processing (Tombaugh & Rees, 2000), also include RT tasks of increasing difficulty and demand (i.e., deciding whether or not a word belongs to four possible semantic categories as quickly as possible). Strong evidence suggests that individuals instructed to simulate brain injury have longer, more variable RTs on these types of tasks than individuals with verified TBI (Reicker, 2008; Willison & Tombaugh, 2006; Woods, Wyma, Yund, & Herron, 2015). Moreover, analysis of RT variables can be used to distinguish individuals feigning impairment from those providing optimal effort (Reicker, 2008; Willison & Tombaugh, 2006)

A number of studies have looked at combining RT with conventional PVTs to improve their specificity and sensitivity. These studies have found that TBI simulators have longer average response times (Bolan et al., 2002; Lupu, Elbaum, Wagner, & Braw, 2018; Rees, Tombaugh, Gansler, & Moczynski, 1998) with increased variability (Lupu et al., 2018; van Hooff, Sargeant, Foster, & Schmand, 2009) compared to healthy adults instructed to provide full effort on forced-choice PVTs. These studies employed analog designs, which are an important step early in the early phases of PVT development (Bianchini et al., 2001); however, without the inclusion of a verified TBI group, the generalizability of findings to clinical practice is unknown. This limitation is even more concerning given the fact that TBI severity (mild vs. severe) appears to moderate classification accuracy when using RT data from more complex RT tasks (Reicker, 2008; Willison & Tombaugh, 2006).

Few studies have investigated patterns of RT on PVTs among groups of clinical interest: individuals with verified TBI and those simulating TBI. Of those studies that exist, mixed results have been found regarding which group displays longer average RTs (Rose, Hall, &

Szaldapetree, 1995; Vagnini, Berry, Clark, & Jiang, 2008). Unfortunately, these studies were subject to many of the previously mentioned limitations common to most analog designs. Namely, these studies did not warn about the presence of PVTs or have test batteries that included tests of neurocognitive functions. Accordingly, their ability to assess the incremental utility of combining RT with standard PVT accuracy to enhance diagnostic accuracy was compromised.

Studies employing known-groups designs have shown that combining time to complete a PVT with traditional PVT accuracy scores improves classification (Erdodi et al., 2017; Kim et al., 2010). However, both studies included mixed clinical samples for which a minority (i.e., less than 10%) of subjects were referred for an assessment related to moderate to severe TBI. Given the evidence that TBI severity moderates RT classification utility, the extent to which findings extend to individuals with TBI of severity exceeding mild was uncertain. Additionally, RT was calculated as the total time for the forced-choice trial and recorded by the examiner's stopwatch, raising concerns over potential measurement error and preventing analysis of more detailed RT characteristics across individual trials.

Not surprisingly, Kanser, Rapport, Bashem, and Hanks (2018) found superior classification accuracies for RT indices when distinguishing TBI simulators and healthy adult comparisons versus distinguishing TBI simulators and individuals with moderate-to-severe TBI. Despite reduced classification accuracy for the comparison of greatest clinical significance, RT indices remained significant individual predictors of group status, added incremental predictive value to traditional TOMM accuracy scores and led to outstanding diagnostic accuracy. Moreover, detailed RT-item characteristics (e.g., ratio of RT for correct to incorrect items) provided initial evidence that RT differences are due, in part, to differences in cognitive

processing. Although RT for incorrect items appear sensitive to this cognitive processing, because most examinees miss very few items, its predictive utility is limited by the lack of accuracy variability on most stand-alone PVTs. Additionally, the common TBI-simulation strategy of deliberate slowed responding mimics a hallmark characteristic of actual TBI (i.e., slowed processing speed), which likely reduces the discrimination power of RT to some extent. Overall, the findings signaled the promise of utilizing biomarkers to enhance diagnostic validity of PVTs; however, it also signaled the need for biometrics that are more sophisticated and specific to the decisional/cognitive processing inherent in TBI simulation than RT.

Section 1.8 – Eye Tracking and Oculomotor Behaviors

Eye movements and various oculomotor behaviors represent a promising covert measure in performance validity assessment. Eye-tracking technologies, the primary method for recording oculomotor behaviors, have seen significant evolution over the past decade. With technological advancements has come increased application to a variety of fields including: neuroscience, advertising/marketing, and driving/aviation training. Eye-tracking applications in these fields have primarily centered on the well-documented finding that oculomotor behaviors provide a quantitative measure of real-time, overt attention (Duchowski, 2002). Specifically, eye fixation on a stimulus serves as biomarker of attention, information processing, and working memory (J. H. Goldberg & Kotval, 1999; Jacob & Karn, 2003; Just & Carpenter, 1976; Orquin & Mueller Loose, 2013). Recording eye movements, therefore, can reveal an individual's patterns of attention over one or more stimuli (e.g., a pictured scene, a written story, etc.) and throughout a given task (e.g., a driving simulation, a recognition memory test, etc.). Specific strategies that occur consciously or unconsciously might be detectable via oculomotor cues. For example, eye-tracking technology could record oculomotor behaviors during PVT performance that may be

sensitive to the common simulation strategy of feigned inattention/poor concentration, especially given that the most successful simulators tend to target visual tasks (Kanser et al., 2017). Specifically, it may be that attempts to appear inattentive entail patterns of visual scanpaths that reveal feigning as compared to natural effort.

In addition to scanpaths recorded by eye-tracking, a wide range of oculomotor behaviors can be reliably recorded (e.g., eye blinks, eye fixations, gaze durations, pupil dilation, saccades, deviations from smooth eye pursuit movements, etc.; Proudfoot, Jenkins, Burgoon, & Nunamaker, 2016). These oculomotor behaviors have been examined in a variety of experimental tasks to enhance our understanding of the cognitive processes involved in language processing (i.e., reading), visual search, and scene perception (see Rayner, 1998, for a comprehensive review). Additionally, research has investigated patterns of various oculomotor behaviors during visual memory tasks. In general, these oculomotor behaviors can be recorded at two levels of analysis. Overall viewing involves recording oculomotor behavior across the entire experimental display. In contrast, directed viewing separates the experimental display into multiple *regions of interest* (ROI) for subsequent analysis. Common oculomotor variables include: number of fixations, proportion of fixations directed to specific ROI, fixation duration (i.e., the length of time the eyes pause on a display), proportion of time in an ROI, and number of transitions in and out of an ROI (Hannula et al., 2010).

Memory research incorporating eye-tracking has shown that analysis of oculomotor behaviors can reveal memory for individual stimuli previously experienced. Research supporting this conclusion has employed a variety of stimuli, including images of famous and non-famous faces (Althoff & Cohen, 1999), familiar and unfamiliar building (Althoff et al., 1998), and novel scenes (Ryan, Althoff, Whitlow, & Cohen, 2000). These studies have revealed that novel items

are viewed with more fixations than pre-experimentally familiar items, and that visual sampling decreases with repeated exposure to pre-experimentally unfamiliar items (Hannula et al., 2010). Moreover, this effect of repetition was observed regardless of task demands (i.e., recognition vs. emotional labeling; (i.e., recognition vs. emotional labeling; Althoff et al., 1998). In addition to memory for individual items, oculomotor behaviors have been shown to index relational memory [i.e., face/scene pairings; (Hannula & Ranganath, 2009; Hannula, Ryan, Tranel, & Cohen, 2007) and temporal relations (Ryan, Moses, & Villate, 2009)].

Most relevant to the present study is the finding that the *novelty preference*, or the increased visual attention to novel items, has been observed during two-item forced-choice recognition tasks (Snyder, Blank, & Marsolek, 2008). Within trials, adults tend to initially look briefly at the target (familiar) stimulus before looking at the foil (unfamiliar) stimulus (Manns, Stark, & Squire, 2000; Snyder et al., 2008). These oculomotor patterns appear to reflect implicit (unconscious) memory more so than explicit (conscious) memory as they occur rapidly and long before explicit behavioral responses are made (Hannula et al., 2010). In sum, oculomotor behavior appears to be guided in consistent, partly obligatory, ways by memory for various aspects of experience (e.g., individual items, relational factors, temporal order; Hannula et al., 2010). Because these oculomotor patterns often occur outside conscious control and awareness, they provide potentially coaching-resistant marker of memory functioning when performance validity is questionable.

Indeed, eye-tracking technology has been effectively used in the area of deception detection, primarily through evidence of concealed recognition (Proudfoot et al., 2016). The concealed information test (CIT) is a recognition-based forensic interviewing test whereby participants view images of crime-relevant and irrelevant images. CITs can present stimuli

sequentially or simultaneously (e.g., four photos in each quadrant of the screen), but generally require respondents to report whether or not they recognize any of the pictured items. Results from analog research designs have shown that individuals instructed to conceal recognition produce oculomotor behaviors that reliably distinguish them from individuals with no prior exposure/recognition instructed to respond honestly (Millen, Hope, Hillstrom, & Vrij, 2017). Consistent with the cognitive load theory of deception (Vrij, Granhag, Mann, & Leal, 2011), which states that deception is cognitively more demanding than honest responding, deceptive responses are related to oculomotor behaviors known to reflect greater depth of processing (i.e., duration and frequency of fixations; (Cook et al., 2012; Griffin & Oppenheimer, 2006). Additional promising eye-tracking measures have included tracking pupil dilation and time spent looking outside ROIs (e.g., the white space between photos) across the administration of a CIT (Proudfoot et al., 2016). Beyond CITs, eye-tracking has been shown to enhance detection of individuals instructed to fake good on a computerized personality measure over and above response extremity and RT variables (van Hooft & Born, 2012). During tasks of relational/associative memory (i.e., face/scene pairings), individuals instructed to simulate memory impairment showed the expected initial, unconscious, automatic oculomotor bias to previously seen targets; however, they spent less time looking at those targets over time compared to healthy adults responding honestly (Mahoney, Kapur, Osmon, & Hannula, 2018). Unfortunately, studies combining eye-tracking data with a traditional PVT used in a clinical setting have not yet been published.

Section 1.9 - Rationale for the Present Study

In contrast to the strong evidence supporting the effect of symptom-specific and test-specific coaching in analog designs, the effects of preparation time and financial incentives on

PVT performance and classification accuracy remain unclear. To date, only one study has empirically assessed the effect of preparation time on simulation performance (Shum, O'Gorman, & Alpar, 2004). Findings from this study are limited by a small TBI sample (n = 15), use of sample-determined PVT cut-off scores, and a preparation condition that involved 1 week to study a handout containing TBI symptom and test-specific coaching information. Participants were not encouraged to use additional resources (e.g., the internet) to prepare for testing. As such, replication and extensions are clearly warranted.

With respect to financial incentives, results have been mixed as to whether they do (Frederick, Sarfaty, Johnston, & Powel, 1994) or do not (Bernard, 1990; Martin, Bolter, Todd, Gouvier, & Niccolls, 1993) affect simulators' test performance and PVT classification accuracy. However, studies reporting no effect either offered small compensation (e.g., $2; Martin et al., 1993) or did not guarantee compensation (i.e., the top two simulators received $50; Bernard, 1990). These differences are important, as relatively small differences in amount of financial incentive (e.g., $40 vs. $15) significantly affect self-reported motivation to successfully simulate (Elhai et al., 2007). Moreover, studies finding no effect of compensation did not provide preparation time, so although they may have been more motivated to successfully simulate, they did not have adequate resources to do so.

Taken together, an ecologically valid simulator group would include the following conditions: (1) symptom-specific and test-specific coaching information; (2) preparation time and encouragement to use additional resources (i.e., the internet); (3) financial incentive that is guaranteed, with opportunity to receive exponentially larger reward for successful simulation; and a test battery that contains both PVTs and tests of neurocognitive functioning. A design containing all of these components would uniquely contribute to the literature concerning the

impact of financial incentives and preparation time on analog designs (a secondary objective of the present study).

As noted, studies combining eye-tracking data with a traditional PVT have not yet been published, although preliminary data from our laboratory (Bashem et al., 2017) and others (Swift, Nicks, Swift, Juan, & Aguerrevere, 2017) have established proof of concept.

Section 1.10 – Aims of the Present Study

The current study seeks to avoid the limitations of prior research by incorporating a unique, ecologically valid analog design to address gaps in the literature concerning the use of biomarkers (e.g., eye-tracking) to distinguish individuals with verified TBI from those instructed to feign neurocognitive impairment. The primary objective of the proposed study is to determine whether eye-tracking can enhance PVT classification accuracy. The *main hypothesis* is that individuals feigning neurocognitive impairment will display unique eye-tracking patterns, and that analysis of these patterns will improve PVT classification accuracy. A secondary objective of the proposed study is to determine the impact of preparation and financial incentives on simulators' test performance and PVT classification accuracy. The *secondary hypothesis* is that the combination of preparation time and incentive will result in a more believable (i.e., successful) simulation. These hypotheses will be tested by the completion of three key objectives.

Objective 1: Compare eye-tracking patterns between full-effort healthy controls, individuals with verified TBI, and TBI simulators on a standard PVT.

Hypothesis 1a. Eye-tracking indices will differ between TBI simulators, individuals with verified TBI and full-effort healthy controls. Candidate indices identified from prior research in our lab (Bashem et al., 2017) include frequency of fixations and fixation transitions, duration of

gaze on correct versus incorrect stimuli (absolute and proportional), and duration of gaze on the background.

Objective 2: Determine the extent to which eye-tracking characteristics provide incremental utility in PVT classification accuracy.

Hypothesis 2a. PVT classification accuracy will improve by combining RT and eye-tracking data with standard accuracy scores.

Objective 3: Determine the extent to which preparation time and financial incentives impact test performance and PVT classification accuracy

Hypothesis 3a. PVT classification accuracy will be better for the simulators in the single-coaching session condition (SIM-SC group) than for simulators in the condition with financial incentive plus preparation time (SIM-IP group).

Hypothesis 3b. Simulators receiving financial incentives and time to prepare (SIM-IP) will have higher rates of successful simulation (i.e., simulators who score impaired on cognitive tests and avoid PVT detection) than simulators who do not receive financial incentive or time to prepare (single-session coaching, SIM-SC).

CHAPTER 2: METHOD

Section 2.1 – Participants

Participants were 163 adults (99 men, 63 women) consisted of three groups of adults: individuals with moderate to severe TBI (TBI Group), healthy comparison adults instructed to put forth best effort (HC Group), and healthy adults instructed to simulate brain injury (SIM-IP Group). The TBI Group (n = 49) 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. Participants in the TBI Group all sustained injuries severe enough to warrant inpatient rehabilitation treatment, > 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 attentional capacity).

Neurologically healthy adults were recruited from the Detroit metropolitan area whose primary language was English and who had no history of neurological conditions (n = 159). Sixty-seven adults were assigned to the healthy comparison group. Forty-seven adults were assigned to a simulator group with a financial incentive and preparation condition (SIM-IP). This SIM-IP group was compared to a simulator group receiving single-session coaching (SIM-SC). Prior to testing, the SIM-SC group was provided information on commons symptoms of TBI and a warning that major symptom exaggeration can be detected; however, they had no financial incentive or time to prepare (n = 45). See *Procedure* for description of the two simulator conditions.

Age of participants ranged from 18 to 78 years. The SIM-IP (M = 33.9, SD = 16.4) group was younger than the HC (M = 46.7, SD = 16.7) and TBI (M = 45.3, SD = 12.7) groups, F (2,

160) = 10.51, $p < .001$. Education ranged from 8 to 20 years. The HC (M =14.1, SD = 2.3) and SIM-IP (M = 14.9, SD = 2.2) groups did not differ significantly with respect to years of education; however, both groups had more years of education than the TBI (M = 12.4, SD = 2.1) group, F (2, 160) = 16.28 $p < .001$.

Section 2.2 Materials and Apparatus

Tobii TX-300 Eye Tracking System. The Tobii TX-300 Eye Tracking System (TX-300) recorded oculomotor behaviors during the administration of all computerized tasks. The TX-300 tracks both eyes using multiple infrared cameras to locate the darkest part of each pupil. Algorithms triangulate the convergence of the eyes' gaze location on the monitor. The TX-300 has a sampling rate of 300 Hz (i.e., one sample every 3 ms). The monitor, a 27" 1080p HD system with a 16:9 aspect ratio, contains the infrared cameras. Gaze data are relayed to Tobii Studio, which assigns the specific coordinate location of the pixel where gaze was directed. The Tobii Studio is also responsible for pre-task calibrations and controlling the start and stop of the TX-300 recording. The apparatus set up was arranged in accordance with Tobii published recommendations, with participants seated approximately 65 cm from the TX-300 monitor during calibration and task administration (See Figure 1; Tobii, 2006).

E-Prime 2.0. A computerized version of the RMT-Words was created in E-Prime 2.0. The orientation of words during stimulus presentation and the forced-choice trial mirrors the traditional paper materials. E-Prime 2.0 recorded response accuracy and reaction time. E-Prime 2.0 runs on its own computer and is linked to the TX-300 using E-Prime Extensions for Tobii (EET; see Figure 2). EET enables E-Prime 2.0 to specify and record oculomotor behaviors of interest. It also provides estimates of the certainty with which accurate data were obtained for oculomotor behaviors in a given 3ms sample for each eye. Tobii and the developers of the

software used to record its data (Psychology Software Tools, Inc., 2009) label this estimate as the "validity" of the eye-tracking data; however, conceptually, it corresponds with measurement precision, which is technically *reliability* in psychometric parlance. Per recommendations of the Tobii manual, only data for which EET was certain that it had recorded all relevant data for at least one eye were used for analyses. Using this standard, Table 1 presents the percentage of participants excluded and retained based on validity cutoffs for all the samples recorded during the task. For example, 94% of the sample had > 50% valid data across the test. A validity cutoff of > 50% was used for all analyses. See Results for further discussion on the criteria for validity.

Section 2.3 – Measures

Participants completed an ecologically valid test battery comprised of embedded PVTs, stand-alone PVTs, and a number of standard tests of neurocognitive functioning. Stand-alone PVTs included computerized versions of the TOMM (Tombaugh, 1996) and the Warrington Recognition Memory Test of Words (described in detail below). The traditional cutoff for Trial 2 of the TOMM was used to classify performance invalidity. The cutoff used for Trial 1 was based on the weighted average of prior research showing various Trial 1 cutoff classification accuracies summarized by Denning (2012). Embedded PVTs included the Reliable Digit Span (RDS) index derived from the Digit Span subtest of the Wechsler Adult Intelligence Scale-Revised (Greiffenstein, 1994) and the Forced-Choice Recall trial of the California Verbal Learning Test-II (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000). With respect to standard tests of neurocognitive functioning, the Wechsler Test of Adult Reading (WTAR; The Psychological Corporation, 2001) was administered as an estimate of overall intellectual ability. Additionally, participants completed the Trail Making Test Parts A and B (Reitan & Wolfson, 1985), the CVLT-II, the Digit Span subtest of the WAIS-IV, and the Symbol Digit Modalities Test (Smith,

1973). Test administration was counterbalanced and all tests were administered and scored according to standard instructions.

Warrington Recognition Memory Test of Words (RMT-Words; Warrington, 1984). The RMT-Words is a 50-item visual recognition test. It consists of a single learning trial in which individuals are presented 50 words. Individuals must state (yes/no) whether they find each word pleasant. Following presentation of the 50 words, individuals complete a single, two-word forced-choice recognition task in which they must select the previously seen word. The RMT-Words was originally developed as a standard measure of recognition memory (Warrington, 1984). Given the fact that it employs a forced-choice paradigm, however, it has become more commonly used as a measure of performance validity (Kim et al., 2010). A large amount of research, employing various research designs (e.g., prevalence, analog, and known-groups designs), has shown the RMT-Words to be an excellent measure of performance validity (H. Goldberg, Back-Madruga, & Boone, 2007; Iverson & Franzen, 1994; Millis, 1992, 1994; Millis & Putnam, 1994; Ross, Putnam, & Adams, 2006; Tardif, Barry, Fox, & Johnstone, 2000). Advantages of the RMT-Words over similar PVTs include shortened time of administration, as it includes only a single learning and forced-choice trial, and the finding that it is only modestly correlated with other PVTs (Nelson et al., 2003).

Oculomotor Behaviors of Interest. Oculomotor behaviors of interest in the present study can be grouped into one of four categories: gazes, fixations, transitions, and durations. A *gaze* occurs whenever the TX-300 detects the location of the eyes on the monitor. A *fixation* is defined by any gaze that is maintained within a particular area of interest (AOI) for at least 150 ms. A *transition* occurs whenever a gaze or fixation is replaced by a new gaze or fixation. In other words, the eyes have moved from one AOI to another AOI. *Durations* are the total time spent

engaged in a particular oculomotor behavior (e.g., a gaze or fixation). Following are the variables of interest examined which were informed by preliminary study findings.

Fixation Variables

Stimulus Trial (Fixation Count-Stim)	Sum of fixations that occurred during the stimulus presentation trial.
Forced-Choice Trial (Fixation Count-FC)	Sum of fixations that occurred during the forced-choice trial.

Transition Variables

Stimulus Trial (Transitions-Stim)	Sum of transitions that occurred across the stimulus presentation trial (i.e., an index of visual scanning).
Forced-Choice Trial (Transitions-FC)	Sum of transitions that occurred during the forced-choice trial (i.e., an index of visual scanning).

Duration Variables

Duration of Fixations within the Background – Stimulus Presentation (Background Fixation – Stim)	The total duration of time spent fixating on the background image during the stimulus presentation.
Duration of Fixations within the Background – Forced-Choice Trials (Background Fixation – FC)	The total duration of time spent fixating on the background image during the forced-choice trials.
Total Duration within the Correct AOI	Total time spent looking (i.e., gaze or fixation) at the correct forced-choice stimulus image (AOI) for the

(Correct Duration)	trial.
Total Duration within the Incorrect AOI (Incorrect Duration)	Total time spent looking (i.e., gaze or fixation) at the incorrect forced-choice stimulus image (AOI) for the trial.
Total Duration of Correct to Incorrect AOI Ratio (Correct/Incorrect Ratio)	Ratio of total time spent looking (i.e. gaze or fixation) at the correct stimulus image to the incorrect stimulus image during the forced-choice trial.

Section 2.4 – Procedure

The TBI Group (n = 49) 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 demographic characteristics including age, gender, years of education, time since injury, and injury severity as assessed via the Glasgow Coma Scale at the time of injury admission to the Emergency Department.

Neurologically healthy participants (n = 159) were recruited from the Detroit Metro 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.), learning disability, serious psychiatric illness, or TBI.

After phone screening, the TBI simulators in the incentive and preparation condition (SIM-IP) were read a scenario over the phone that described their involvement in litigation for a TBI they sustained following a motor vehicle accident. The core script from this scenario has

been used successfully in TBI simulation studies with similar research designs (Tombaugh, 1997). SIM-IP participants were then read a common set of symptoms that can occur following TBI (slowed thinking, memory dysfunction, etc.) and warned about the presence of PVTs within the test battery. They were informed that if they were able to appear brain injured successfully, while avoiding PVT detection, they would receive an additional $30 and be entered into a raffle to win $300. Finally, they were encouraged to use resources such as the internet to help prepare them for their testing session, which was scheduled within 1 to 2 weeks from the initial phone screening. This scripted information (See Appendix A) was subsequently emailed to SIM-IP participants. The SIM-SC participants (no-incentive/no-preparation condition) were collected during preliminary studies; therefore, there was no conflict regarding inequitable compensation between the groups.

To obtain an estimate of general intelligence, all participants were administered the WTAR under standard conditions, with instruction to perform to the best of their abilities. The TBI Group (n = 49) and HC Group (n = 67) completed the remainder of the battery under standard instruction to put forth best effort. The SIM-IP Group was informed that their performance on the WTAR was separate from the TBI simulation (i.e., it is important they do not attempt to simulate TBI on this test) and would not affect their opportunity to win the added financial incentive.

Following completion of the WTAR, the SIM-SC Group and SIM-IP Group were read the same scenario that describes their involvement in litigation for a TBI they sustained following a motor vehicle accident. To summarize, participants were instructed to pretend they are seeking compensation after sustaining a TBI in a motor vehicle accident. It is explained that although they do not have any real current symptoms, their goal is to appear brain damaged in

order to obtain a larger settlement. They were warned about presenting believable deficits and the presence of PVTs within the test battery. They were given time to read a pamphlet that describes common symptoms that can occur following TBI, and reminded of their opportunity for additional compensation. The SIM-IP Group and SIM-SC Group then completed the remainder of the assessment battery under instructions to feign TBI.

Following completion of the test battery, all simulators were given a questionnaire containing open- and closed-ended questions pertaining to their simulation approach. Questions examined whether they attempted to simulate TBI as instructed and the strategies they employed. Additionally, they were asked whether they believed each test in the battery was a test of effort, neurocognitive functioning, or both (Tan et al., 2002). SIM participants from both groups (SIM-SC and SIM-IP) were excluded from analysis if they report not attempting to simulate 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.

Section 2.5 – 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, and collinearity). *Descriptive statistics* were conducted to describe the sample demographics and PVT performance. To examine demographic differences, the groups were compared on age, years of education, and gender proportions using one-way analysis of variance (ANOVA) and χ^2 as appropriate.

Hypotheses 1a was tested using Kruskal-Wallis tests with group (TBI, HC, SIM-IP), as the between-subjects factor and eye-tracking indices as the within-subjects factor. Post hoc comparisons were performed using Mann-Whitney tests as appropriate.

Hypothesis 2a was addressed using multivariable binary logistic regression, testing the individual and combined predictive values of the traditional accuracy index and various eye-tracking indices. Logistic regression models with group membership (HC vs. SIM-IP; TBI vs. SIM-IP) as the outcome variable were fitted for promising eye-tracking indices separately (i.e., single-variable models for those showing significant Mann-Whitney comparisons). Multivariable logistic regression models combined PVT accuracy and eye-tracking variables as covariates, with group membership (HC vs. SIM-IP; TBI vs. SIM-IP) as the outcome variable. The combined models were examined to determine the extent to which eye-tracking data improved model classification over standard scoring (i.e., number correct). The diagnostic accuracy of these models were further assessed through analysis of receiver operating characteristics (ROC).

Hypothesis 3a was tested by comparing PVT failure rates of SIM-IP and SIM-SC. Hypothesis 3b was tested by examining group differences (SIM-IP vs. SIM-SC) in rates of successful simulation.

CHAPTER 3: RESULTS

Validity

Table 1 presents the percentage of participants excluded and Spearman rho correlations of eye-tracking indices with RMT-Words accuracy as a function of different validity cut scores. Not surprisingly, as validity cutoffs gradually increase (i.e., reliability increases) the strength of correlations gradually increase. Table 2 shows group differences in the frequency of excluded cases across validity cut scores. The number of participants excluded at a validity cut score of $\geq$ 50% was not significantly different across groups. As validity cut scores increased, the number of participants excluded did differ significantly across groups. More specifically, the TBI group was disproportionately excluded as validity requirements increased. The 50% cutoff was used for all subsequent analyses in an effort to maximize ecological validity, minimize disproportionate exclusion of the TBI group, and adopt the most conservative statistical approach.

Intercorrelations

Table 3 presents descriptive correlations for RMT-Words indices and demographic variables. Notably, none of the eye-tracking indices were redundant with the traditional RMT-Words accuracy score. Medium to large correlations were seen between RMT-Words accuracy and all eye-tracking indices recorded during the *forced-choice* trials (*r*s -.34 to -.54). None of the corresponding eye-tracking indices recorded during the *stimulus trials* had meaningful correlations with RMT-Words accuracy (*r*s .01 to -.16). Not surprisingly, forced-choice duration variables related to decision-making (e.g., Correct Duration, Incorrect Duration) were strongly correlated with one another (r = .92) as well as to the number of fixations occurring across forced-choice trials (e.g., Fixation Count-FC; *r*s .78 to .83). In other words, participants who had more fixations (i.e., looked back and forth between the correct and incorrect response options)

spent more time looking at the correct and incorrect response options. In contrast, the forced-choice duration variable unrelated to decision-making (BackgroundFix-FC; r = .70) was strongly correlated with the number of transitions (i.e., frequency of changes in gaze across the screen) and only weakly correlated with the number of fixations (r = .21). It is important to note that although groups differed with respect to demographic variables, none of the demographic variables had meaningful correlations with RMT-Words accuracy or eye-tracking indices (all r < .20).

Table 4a presents descriptive statistics and univariate group comparisons of RMT-Words indices. Kruskal-Wallis tests revealed a number of eye-tracking indices that differed significantly across groups. Significant group differences were observed across *all* eye-tracking variables recorded during *forced-choice trials*. *None* of the eye-tracking variables recorded during the *stimulus trials* differed significantly across groups, and all group contrasts for the stimulus trials showed small effect sizes (Cohen's d < 0.3, mean and median d = 0.1). Table 4b presents Mann-Whitney post hoc contrasts and effect sizes across pairwise group comparisons. With the exception of Background Fixation-FC comparing TBI and SIM-IP, all eye-tracking variables recorded during the forced-choice trials differed significantly across the comparisons of greatest interest (i.e., SIM-IP vs. TBI and SIM-IP vs. HC) with medium (d > .50) to large (d > .80) effect sizes (mean d = 0.9, median d = 1.0). In general, larger effect sizes were observed between the SIM-IP and HC groups in comparison to those observed between the SIM-IP and TBI groups. To summarize, the SIM-IP group answered significantly fewer items correctly and had significantly more transitions, fixations, and time spent looking at the response options compared to the HC and TBI groups. The TBI group answered significantly fewer items correctly than the HC group. Notably, the TBI and HC group did not differ significantly with respect to time spent looking at

the response options (i.e., Correct Duration, Incorrect Duration, Correct/Incorrect Ratio Duration), suggesting that these indices may be more robust to the effects of cognitive impairment than traditional accuracy scores.

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

Table 5a presents classification accuracy statistics for the RMT-Words indices as predictors of group status. Total correct (RMT-C) showed the best balance of hit rate (83%), sensitivity (70%), and specificity (93%). The highest specificities (100%) were observed for eye-tracking indices recorded during the stimulus trials; however, this is because all participants were predicted to be HC. Of the eye-tracking indices, Fixation Count-FC and Incorrect Duration showed the best balance of hit rate, sensitivity, and specificity.

Logistic regression statistics are presented in Table 5b. All eye-tracking indices recorded during the forced-choice trials were significant individual predictors of group status. In contrast, Background Fixation-Stim was the only eye-tracking index recorded during the stimulus trials that was a significant predictor. Area under the receiver operating characteristics (ROC) curve values were calculated as a means of quantifying the discriminability of these models (Table 5a). Traditional cutoffs classify AUC values as "acceptable" ($.70 \leq AUC \leq .79$), "excellent" ($.80 \leq AUC \leq .89$), or "outstanding" ($AUC \geq .90$; Hosmer & Lemeshow, 2000). The only variable showing "outstanding" discrimination was the traditional accuracy score (RMT-C). The following eye-tracking variables showed "excellent" discriminability: Transitions-FC, Fixation Count-FC, Background Fixation-FC, and Incorrect Duration. "Acceptable" discrimination was observed for Correct Duration and Correct/Incorrect Ratio. HC participants made very few errors on the RMT-Words; however, the error rate and variability were notably greater on the RMT ($M = 48.8$, $SD = 2.1$) than has been observed for these same HC participants on the TOMM (M =

49.7, SD = 0.7; Kanser et al., 2018). Accordingly, two-variable predictive models displayed adequate stability and could be explored.

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

Tables 5a and 5b show classification and model fit statistics for the hierarchical two-variable logistic regression models predicting group membership (SIM-IP vs. HC). All of the two-variable models combined RMT-C (Block 1) with an eye-tracking index (Block 2) that was a significant individual predictor of group status. Although all of the two-variable models were significant ($p < .001$), not all eye-tracking indices added *incremental* predictive value (i.e., $p < .05$ in Block 2) to RMT-C. Eye-tracking indices adding incremental predictive value included: Transitions-FC, Fixation Count-FC, and Background Fixation-FC. ROC curve analyses showed that the addition of eye-tracking indices led to minimal improvements in the RMT-C's "outstanding" discrimination (i.e., AUC increases ranging .01 to .02).

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

Table 6a presents classification accuracy statistics for the RMT-Words indices as individual predictors of group status (SIM-IP vs. TBI). In contrast to the comparisons with HC, RMT-Words accuracy (RMT-C) was *not* the best predictor of group status. Eye-tracking variables exceeding RMT-C in terms of classification accuracy included Incorrect Duration (Hit Rate = 72%, Sensitivity = 64%, Specificity = 80%) and Fixation Count-FC (Hit Rate = 71%, Sensitivity = 60%, Specificity = 82%).

Table 6b shows the single-variable models that were significant ($p < .05$) predictors of group membership. Relative to the comparisons with HC, fewer RMT-Words indices were significant individual predictors of group status, and discriminability of eye-tracking indices generally surpassed that of traditional accuracy scores. Significant predictors included: RMT-C,

Fixation Count-FC, Correct Duration, Incorrect Duration, and Correct/Incorrect Ratio. All of the indices that were significant predictors displayed "acceptable" discriminability, with the exception of Correct Duration, which had an AUC just outside the "acceptable" range (AUC = .68, 95% C.I. = 0.57-.079).

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

Tables 6a and 6b show classification and model fit statistics for the hierarchical two-variable logistic regression models predicting group membership. Again, all of the two-variable models combined RMT-C (Block 1) with an eye-tracking index that was a significant individual predictor of group status (Block 2). Table 6b shows that each of those eye-tracking indices (Fixation Count-FC, Correct Duration, Incorrect Duration, and Correct/Incorrect Ratio) added significant incremental predictive value to RMT-C.

ROC curve analyses revealed that AUC of all the two-variable models improved upon the single-variable model using only RMT-Words accuracy (AUC = .75). With the exception of Correct Duration, the addition of eye-tracking indices in the two-variable models led to "excellent" discriminability (AUC $\geq$ .80). The leading two-variable model in terms of AUC was Incorrect Duration combined with the traditional RMT-C accuracy score (AUC = .81). Of the eye-tracking indices that added incremental predictive value to RMT-C, AUC increases ranged from .03 (RMT-C + Correct Duration, AUC = .78) to .06 (RMT-C + Incorrect Duration, AUC = .81).

Rates of Successful Simulation (SIM-IP vs. SIM-SC)

The series of analyses testing the influence of financial incentive and preparation time on likelihood of successful simulation included all participants with valid data on the PVTs and neuropsychological battery, regardless of whether they produced valid eye-tracking data. Table

7a presents group comparisons of demographic variables, standardized neuropsychological test performances, and rates of impaired performances across neuropsychological tests. There were no significant differences between SIM-IP (n = 49) and SIM-SC (n = 46) with respect to education or estimates of intellectual functioning. The SIM-IP group was significantly younger (p = .005) than the SIM-SC group. This difference in age showed a medium effect size (Cohen's d = 0.58). Age is not considered an important confounding variable because neuropsychological test performance was compared using age-adjusted T scores. Moreover, the PVTs listed in Table 7b have been shown to be resistant to the effects of age (Constantinou & McCaffrey, 2003; Teichner & Wagner, 2004).

With respect to neuropsychological test performance, SIM-IP performed significantly worse than SIM-SC on Trails B, with a medium effect size (d = 0.64). Though not statistically significant, small but nontrivial effect sizes were observed for the group contrast on Trails A (d = 0.33, p = .052) and total words recalled across Trials 1-5 of the CVLT-II (d = 0.24, p = .125). Neuropsychological test performance was statistically equivalent across the remaining indices with negligible effect sizes (mean d = 0.09). Frequency analysis of impaired scores on the neuropsychological tests shows that SIM-IP were significantly more likely than SIM-SC to score > 1 SD below the mean on Trails A (ϕ = .21, p = .045) and Trails B (ϕ = .24, p = .019). Using a more stringent cutoff to define impairment, > 1.5 SD below the mean, SIM-IP was significantly more likely to produce impaired Trails B performance (ϕ = .25, p = .015) than SIM-SC. Groups did not differ significantly with respect to rates of impaired performance across the remaining neuropsychological tests using either cutoff.

Table 7b presents group differences in PVT performance, rates of PVT failure, and rates of successful simulation. Successful simulation was defined by failure on 0-1 PVTs as well as at

least one impaired neuropsychological test performance. With respect to PVT performance, SIM-IP had significantly greater accuracy performance than SIM-SC on Trial 1 of the TOMM, with a small effect size (d = 0.38). Small but nontrivial effect sizes were also observed across TOMM Trial 2 (d = 0.24) and RMT-C (d = 0.24), as SIM-IP had greater accuracy performance than SIM-SC; however, group differences were not statistically significant. Frequency analysis of PVT failure revealed that SIM-IP were significantly less likely than SIM-SC to fail Trial 1 of the TOMM (ϕ = .24, p = .021). Rates of successful simulation did not differ across groups using liberal (> 1 SD below the mean) and conservative (> 1.5 SD below the mean) cutoffs to define impaired neuropsychological test performance.

Exploratory analyses tested the utility of eye-tracking biometrics to enhance detection among the combined subgroup of successful simulators in either the SIM-SC or SIM-IP groups (n = 37 with viable eye-tracking data). In this rigorous test, Mann-Whitney tests indicated that several eye-tracking variables continued to show significant differences from healthy comparisons (frequencies of transitions and fixations, durations of gaze to correct and incorrect targets, and ratio of gaze to correct vs. incorrect targets), with medium (d = 0.5, Correct duration) to large (d = 1.1, transitions) effects. Similarly, Mann-Whitney tests indicated that several eye-tracking variables continued to show significant differences from individuals with TBI (frequencies of fixations, durations of gaze to correct and incorrect targets), with medium (d = 0.4 to 0.6) effects. For the comparison of successful simulators to TBI, logistic regressions indicated improvements in classification (i.e., additional cases of simulators detected) of 32.4% (fixations), 21.6% (correct duration), and 29.7% (incorrect duration).

CHAPTER 4: DISCUSSION

Findings provide support for the hypothesis that oculomotor behaviors can serve as covert biomarkers sensitive to feigned neurocognitive impairment. A number of eye-tracking indices showed excellent discrimination between genuine performance and feigned neurocognitive impairment that, in some instances, exceeded the discriminability of a stand-alone performance validity test, the Warrington Recognition Memory Test of Words (RMT-Words). Moreover, findings demonstrated that these oculomotor behaviors tap unique aspects of feigned neurocognitive impairment (i.e., beyond overt accuracy responses) and, therefore, can enhance the classification accuracy of traditional performance validity tests. Eye-tracking indices added incremental predictive value to the traditional RMT-Words accuracy score, leading to excellent diagnostic accuracy. The extent to which oculomotor behaviors surpassed and added incremental predictive value to the traditional RMT-Words depended on the groups being compared. In contrast to the strong support for the central hypothesis regarding the eye-tracking utility, findings provide mixed support for the hypothesis that preparation time and financial incentives facilitate a more believable and successful TBI simulation. Although these factors appeared to affect performance, individuals provided with these enhancements were not more likely to accomplish the challenging goal of performing impaired on tests of cognitive ability while at the same time avoiding PVT detection.

Objective 1: *Compare eye-tracking patterns across groups*

Investigated oculomotor behaviors were initially selected based on their established construct validity as biomarkers of relevant cognitive processes (i.e., attention, working memory, information processing; Hannula et al., 2010; Orquin & Mueller Loose, 2013). Additionally, eye-tracking index selection was influenced by preliminary research from our lab establishing proof

of concept (Bashem et al., 2017) as well as findings from research investigating oculomotor patterns of feigned memory associations and deception (Millen et al., 2017; Proudfoot et al., 2016).

The relationships between eye-tracking variables and other variables of interest was investigated first across the entire sample. Overall, the pattern of relationships among oculomotor behaviors and other variables of interested varied with task demands as predicted by theory. Stimulus trials require basic, sustained attention. In the absence of signficant off-task behaviors (i.e., frequently looking off screen, not directing gaze to the stimuli), oculomotor behavior variability is reduced and should not be related to performance accuracy. Indeed, none of the oculomotor behaviors recorded during stimulus trials showed meaningful relationships with performance accuracy. In contrast, forced-choice trials of the RMT-Words require basic attention, information processing, and recognition memory. Not surprisingly, oculomotor behaviors previously shown to index these cognitive abilities showed strong relationships with RMT-Words accuracy performance during forced-choice trials. Consistent with research showing that visual sampling decreases with increased familiarity (Hannula et al., 2010), performance accuracy increased as the number of fixations and time spent looking at response options decreased. In other words, participants spent less time fixating on response options the more certain and familiar they were with one of the forced-choice items. Similarly, more scanning behaviors and time spent fixating on the background during forced-choice trials was associated with decreased performance accuracy.

A similar pattern emerged when comparing oculomotor behaviors across groups of interest. Oculomotor behaviors were relatively similar across groups during stimulus trials. In contrast, *all oculomotor behaviors showed notable differences across groups during forced-*

choice trials. Consistent with previously published research in deception, individuals feigning neurocognitive impairment had more fixations and spent more time looking at response options than did healthy adults responding honestly (Cook et al., 2012; Griffin & Oppenheimer, 2006). Individuals feigning impairment also engaged in more oculomotor behaviors unrelated to active decision-making (i.e., scanning behaviors and time spent looking at the background). This latter finding may be the result of conscious attempts to feign inattention and/or slowed processing speed, two of the most commonly used strategies of individuals instructed to feign TBI (Kanser et al., 2017; Tan et al., 2002).

Consistent with preliminary findings from our lab using a similar forced-choice PVT, individuals feigning neurocognitive impairment engaged in more fixations, scanning of the screen, time spent looking at the background of the screen, and time spent looking at response options than did healthy adults and individuals with verified TBI (Bashem et al., 2017). Similarly, individuals with verified TBI engaged in more fixations, scanning behavior, and time spent looking at the background of the screen compared to healthy adults. Findings suggest the differences in oculomotor behaviors unassociated with active decision-making (e.g., frequency of scanning behavior and time looking at the background) between healthy adults and those with TBI may be the result of genuine cognitive inefficiency and inattention. Although healthy adults and individuals with verified TBI spent a similar amount of time looking at correct and incorrect response options (i.e., gaze duration), those with TBI looked back and forth between response options more often than their healthy counterparts. This larger number of fixations likely reflects greater indecision/uncertainty across items and an associated increased depth of processing. Consistent with this interpretation, individuals with verified TBI did commit significantly more errors than healthy adults. It is noteworthy that the TBI group committed more errors on the

RMT task than did healthy adults, because research on the TOMM – the most widely used PVT – generally reports a ceiling effect and equivalent accuracy (Kanser et al., 2017; Rees et al., 1998; Sollman & Berry, 2011; Tombaugh, 1996, 1997). Because the RMT was created as an actual clinical test (Warrington, 1984), it is more sensitive to cognitive impairment than a typical PVT and yielded a broader range of scores. The broader range of scores poses an advantage for the RMT over other tasks used as PVTs, both for use in clinical applications and also for statistical issues involving viable model stability. It is also noteworthy that the general effect of scanning behavior during forced-choice trials was large, whereas the effect of fixations was medium. The latter is more specific to decisional process. In other words, the effect of TBI on organizing and directing attention during the task is larger than the effect of decisional effort in comparison to healthy adults providing full effort. Notably, negligible effects were observed for the ratio of time spent looking at the correct to incorrect response option, which inherently accounts for the slowed processing speed and possibly greater indecision of those with TBI compared to HC.

The fact that adults with verified TBI committed significantly more errors but spent similar amounts of time looking at response options compared to healthy adults suggest that the latter index (gaze duration) is more robust to the effects of TBI. In contrast, gaze duration for correct and incorrect response options was very effective at differentiating TBI simulators from the two groups providing full effort (healthy adults and verified TBI), with generally large effect. Moreover, this differential pattern of group differences that dissociates response accuracy from gaze duration suggests these oculomotor behaviors are sensitive to the unique, increased cognitive demand inherent in deception/simulation (Vrij et al., 2011). Individuals feigning neurocognitive impairment must not only identify the correct response option, but also decide

whether or not to respond correctly. This added cognitive processing may unconsciously bias attention, and therefore oculomotor behaviors, towards incorrect response options. In contrast, individuals responding honestly tend to spend less time looking at novel features (i.e., foils) after they have recognized the correct response option (Hannula et al., 2010). Consistent with this interpretation, oculomotor behaviors incorporating time spent looking at incorrect response options (e.g., incorect duration, correct/incorrect duration ratio) were notably different for individuals feigning impairment compared to those responding honestly.

Objective 2: *Determine the incremental utility of combining eye-tracking indices with RMT-Words accuracy*

To date, studies combining eye-tracking data with a traditional PVT have not yet been published, although preliminary data from our laboratory (Bashem et al., 2017) and others (Swift et al., 2017) have established proof of concept. Findings provide support for the central hypothesis that combining oculomotor behaviors recorded through eye-tracking technology with RMT-Words accuracy can enhance its diagnostic accuracy. Not only were a number of oculomotor behaviors excellent individual predictors of performance validity, but they also added incremental predictive value to the RMT-Words accuracy. The strength of these findings varied according to what populations were being compared.

Healthy adults versus simulators. When comparing individuals without genuine neurocognitive impairment (SIM-IP vs. HC), RMT-Words accuracy was the best single predictor of group membership, showing "outstanding" group discrimination. This finding is consistent with the large body of research that has consistently shown the RMT-Words to be an excellent measure of performance validity (H. Goldberg et al., 2007; Iverson & Franzen, 1994; Millis, 1992, 1994; Millis & Putnam, 1994; Ross et al., 2006; Tardif et al., 2000). All oculomotor

behaviors recorded during the forced-choice trials were significant predictors of group membership. Time spent fixating on the background was the only significant predictor of group status recorded during stimulus trials. Healthy adults spent more time fixating on the background than adults feigning impairment. The effect size was small, but the finding is consistent with the pattern indicating that simulators focus intensively on gazing at the stimuli relative to expending cognitive resources on the background during the critical learning phase. Consistent with prior research (Proudfoot et al., 2016), adults feigning impairment spent more time fixating on the background than HC during forced-choice trials. During forced-choice trials, the number of fixations and time spent looking at the incorrect response option were the best predictors among the eye-tracking variables to discriminate healthy adults from simulators, showing "excellent" discriminability.

Adults with TBI versus simulators. As in prior research, indexes that showed powerful ability to discriminate between healthy adults and simulators in the typical analogue design were less powerful in the ecologically valid challenge of discriminating the groups of greatest clinical interest: individuals feigning TBI and those with verified TBI. Oculomotor behaviors partly under conscious control (i.e., purposefully looking at the background instead of at the response options) no longer reliably distinguished groups, suggesting TBI simulators were able to adequately feign some aspects of inattention/cognitive inefficiency similar to adults with verified TBI. Although the magnitude of the effect was not as strong as that observed for comparisons of healthy adults and simulators, indexes involved in decision-making and perhaps largely outside of conscious control (number of fixations and relative time spent looking at response options) also discriminated simulators from adults with verified TBI. Similarly, the RMT-Words accuracy showed less powerful classification accuracy when distinguishing performance invalidity from

genuine neurocognitive impairment (SIM-IP vs. TBI). This finding is consistent with research showing reduced RMT-Words performance in populations with genuine neurocognitive deficits as well as its origin as a task of neurocognitive ability (Kim et al., 2010; Warrington, 1984). The performance accuracy of these individuals with verified TBI was consistent with those observed in the few other studies examining groups comprised of primarily moderate to severe TBI (Iverson & Franzen, 1994, 1998; Millis, 1992, 1994). Notably, *oculomotor behaviors were better able to distinguish feigned and genuine neurocognitive impairment than RMT-Words accuracy*. More specifically, the number of fixations, time spent looking at the incorrect response option, and ratio of time spent looking at correct to incorrect response options were better predictors of group membership than RMT-Words accuracy. Moreover, each of these four indexes showed incremental utility to the traditional accuracy score in combined predictive models.

Because no formal standard exists for evaluating enhancements in diagnostic accuracy, the current study relied on the investigating changes in several indicators of classification. Specifically, the current study subjectively weighed changes observed in hit rate, sensitivity, specificity, and AUC values. AUC was ranked as the most important statistic because it is an objective measure that considers the ranges of sensitivities and specificities in the sample. Following AUC, interpretative value was placed on changes in hit rate, specificity, and sensitivity.

Overall, findings provide support for the central hypothesis that combining eye-tracking data with RMT-Words accuracy could enhance its diagnostic accuracy. For the comparison of greatest clinical interest (SIM-IP vs. TBI), all the oculomotor behaviors associated with active decision-making added incremental predictive value and led to “excellent” discrimination. In comparing healthy adults and those feigning neurocognitive impairment, the potential for

improvements in diagnostic accuracy was limited by the fact that RMT-Words accuracy showed "outstanding" discrimination and made few errors in classification. Despite this challenge, number of fixations, frequency of scanning behaviors, and time spent looking at the background during forced-choice trials added unique predictive value in discriminating simulators from healthy adults.

***Objective 3**: Determine the extent to which preparation time and financial incentives affect test performance and PVT classification accuracy*

Findings add to the limited research investigating the effects of experimental manipulations designed to enhance the generalizability of simulator groups to clinical malingerers. With respect to financial incentives, previous investigations have revealed mixed effects that were generally limited by financial incentives that were small or not guaranteed, test batteries that included only PVTs, and not providing simulators time to prepare (Bernard, 1990; Frederick et al., 1994; Martin et al., 1993). Overall, the combined effect of preparation time and financial incentive led to a pattern of neuropsychological test performance that was more consistent with patterns observed in TBI. Specifically, individuals with preparation time and incentive displayed more consistent and significant impairment across tasks of processing speed and executive functioning, cognitive domains particularly sensitive to the effects of TBI (Axelrod, Fichtenberg, Liethen, Czarnota, & Stucky, 2001, 2002; Cicerone et al., 2011; Dikmen et al., 2009).

Enhanced coaching and classification accuracy. With respect to Hypothesis 3a, individuals receiving preparation time and incentive were more likely to avoid detection on one of five PVT indices; additionally, they performed somewhat better than did those who did not receive preparation time or financial incentive across three of the five PVT indices, although the

effect was small. These findings are consistent with the one study, to date, that has looked at the effect of preparation time and financial incentive (Shum et al., 2004). Preparation time and financial incentive were once again found to affect performances on tests of neurocognitive ability and PVTs: these participants performed equivalent to participants without enhanced preparation on all cognitive tests, except Trails A and Trails B, on which they performed worse and more frequently met criteria for "impairment." It is noteworthy that a pattern of *specific* impairment emerged, and especially that performance on most tests was equivalent, given that the operational definition of the group included a requirement to perform > 1 standard deviation below the normative mean on at least one cognitive test. The observation of a specific pattern among the successful simulators suggests that they took a targeted approach. Thus, the experimental manipulation appeared to work in the direction predicted by theory; however, the magnitude of the effect did not translate into meaningful differences in classification accuracy across most PVTs.

Both simulator groups in the current study received test and symptom-specific coaching, manipulations that have consistently been shown to affect PVT performance and oftentimes, classification accuracy (Brennan et al., 2009; Suhr & Gunstad, 2007). It appears as though the provision of financial incentive and preparation time add, at most, minimal effects above and beyond the combined effect of test and symptom-specific coaching. Simulators were encouraged to prepare using the internet and even provided key terms to help guide their search. Although information that threatens PVT security is readily available on the internet (Bauer & McCaffrey, 2006; Ruiz et al., 2002), it may be that this information cannot be successfully identified, interpreted, or utilized by the average adult with limited exposure to psychological assessment/psychometrics. Additionally, it may be that the financial incentive in the current

study did not motivate participants to spend enough time searching the internet in preparation. However, qualitatively, participants in the condition reported that they did search the internet prior to the experiment.

Enhanced coaching and successful simulation. Findings do not support Hypothesis 3b. The provision of preparation time and financial incentive did not lead to higher rates of successful simulation. In other words, individuals receiving incentive and preparation time were not more likely to accomplish the challenging goal of performing impaired on tests of cognitive ability while at the same time avoiding PVT detection. Similarly, prior research has shown that although experimental manipulations (e.g., increasing levels of test-specific coaching) can affect PVT performance and classification accuracy, simulators will generally still fail at least two or more PVTs when completing a neuropsychological battery containing stand-alone and embedded PVTs (Lau et al., 2017).

Overall, the rate of successful simulation in both groups was approximately 40% with a liberal criterion, expecting no more than one PVT failure and an impairment defined as performance at one standard deviation below the normative mean or greater. At a more rigorous criterion defining impairment (1.5 standard deviations below the normative mean), successful simulation approached 33 – 35%.

Limitations

The most notable limitation of the present study was the fact that groups differed with respect to the validity of their obtained eye-tracking data. In other words, at a rigorous criterion for defining reliable capture, the eye-tracking system was less certain where some adults with TBI were looking throughout the task as compared to healthy adults and those feigning neurocognitive impairment. This limitation is not surprising when one considers TBI is

associated with long-standing oculomotor deficits that may affect eye-tracking validity, leading to greater uncertainty in gaze location (Mani, Asper, & Khuu, 2018). Moreover, all TBI participants completed the RMT-Words in the context of a complete neuropsychological battery that required approximately 2 to 2.5 hours to complete. As such, the factors associated with TBI that generally impact the validity of standard neuropsychological tests and evaluations (e.g., fatigue, irritability, inattention) could have contributed to the lower rates of eye-tracking reliability in the TBI group. The present study managed the issue of differential reliability by choosing the lowest acceptable criterion that yielded equivalent exclusions among the groups.

The present study elected to include participants for whom at least 50% of gazedata was deemed valid by the eye-tracking system. Unfortunately, no standard validity criteria cutoffs exist and studies must choose whether to place greater emphasis on inclusion/external validity or exclusion/internal validity. Using a cutoff of 50% validity likely diminished the strength of findings by increasing within-group variability across eye-tracking indices; therefore, the present findings reflect a conservative estimate of the power of eye-tracking biometrics in this context. However, using a cutoff of 80-90% would have led to differential exclusion of the TBI group, significantly reducing the ecological validity of the study and limiting findings to a subset of individuals with moderate-to-severe that can produce near perfect eye-tracking data. Regardless, the issue of eye-tracking validity in the context of TBI appears to be a concern inherent in the clinical condition as opposed to an idiosyncratic limitation of the current study. This issue of differential reliability and potential for selection bias should be investigated in future research.

Another limitation of the present study is the generalizability of the sample. The TBI group consisted of individuals with well-documented histories of moderate to severe TBI. It is unclear the extent to which findings may generalize to individuals with mild or mild complicated

TBI. Concern regarding the generalizability of analog design simulator groups has been well discussed and explored (G. J. Larrabee, Greiffenstein, Greve, & Bianchini, 2007; Rogers, 1997; Suhr & Gunstad, 2007). Although the current study offered financial incentives, they pale in comparison to the real world financial incentives available in clinical contexts. Moreover, individuals feigning neurocognitive impairment in clinical contexts are likely not motivated solely by financial incentives. There exist a wide range of primary and secondary gains that may consciously and unconsciously impact test performance. These added elements were not tested in the present study. The simulator group was also given at least 1 week to prepare as well as symptom and test-specific coaching; however, individuals in clinical and forensic contexts likely have more time to prepare for their neuropsychological evaluations, and their preparation may be monitored and boosted by unscrupulous attorneys. Although coaching about detection of malingering and test content are common, the extent and depth of test-specific coaching informed attorneys typically provide is unclear (Essig et al., 2001). The current study provided general coaching strategies; however, research suggests that depth of test-specific coaching can significantly impact PVT performance and classification (Lau et al., 2017).

Conclusions and Future Directions

Overall, given the combined evidence from this study, it seems reasonable to suggest that it is moderately difficult to fake moderate-to-severe TBI, even under conditions of explicit and sophisticated coaching, among healthy adults with ample time to prepare, who are motivated is succeed at doing so. In this study, 35-40% were successful under the optimal conditions to fake successfully, depending on the rigor of the criterion for successful simulation. This study provides strong support for the use of eye-tracking technologies in the assessment of performance validity as a means to enhance detection of feigned impairment. Oculomotor

behaviors show considerable promise as covert biomarkers sensitive to the unique cognitive processes inherent in deception and feigned neurocognitive impairment. Moreover, these oculomotor behaviors added incremental predictive value to a stand-alone PVT, suggesting they are sensitive to aspects of feigned impairment that are unaccounted for by stand-alone PVTs. Lastly, to our knowledge, this is the first study to show that these covert biomarkers can exceed PVT classification accuracy in distinguishing feigned neurocognitive impairment from verified TBI. In sum, combining eye-tracking with stand-alone PVTs appears to be an effective, yet currently expensive, strategy for enhancing the assessment of performance validity without prolonging the time of neuropsychological evaluations.

It is important to reiterate the fact that findings varied according to what groups were being compared (i.e., healthy adults providing full effort or adults with verified TBI), highlighting the importance of including individuals with the clinical syndrome of interest in studies utilizing analog designs. The differential rates of validity in eye-tracking data only further emphasize this point. Eye movement abnormalities associated with TBI likely contributed to reduced eye-tracking validity; however, TBI is not unique in this regard. Eye movement abnormalities have also been demonstrated in a variety of other neurologic (Antoniades & Kennard, 2015) and psychiatric conditions (Bittencourt et al., 2013). As such, replication of the current study with other groups of clinical interest is strongly recommended.

Findings from the current study provide limited support for the incremental value of preparation time and financial incentive to the well documented effects of symptom and test-specific coaching in analog designs. Research investigating the effects of various degrees of test-specific coaching has shown promise and warrants future replication (Lau et al., 2017).

APPENDIX A (SCRIPT)

In this study you will be asked to complete two tasks that are often used to measure a variety of changes that occur in people who have brain damage. As you take each test, I would like you to assume the role of someone who has experienced some brain damage from a car accident.

Pretend that you were involved in a head-on collision. You hit your head against the windshield and were unconscious for 15 minutes. You were hospitalized overnight for observation and then released. Gradually, over the past few months, you have started to feel normal again. However, your lawyer has informed you that you may obtain a larger settlement from the court if you look like you are still suffering from brain damage.

As you portray the above person, try to approach each test as you imagine this person would respond if the individual had been given the same instructions from his or her lawyer. Perform on the tests in such a way as to convince the examiner that you are truly brain damaged, keeping in mind that settlement monies depend upon your being diagnosed as cognitively impaired.

Also be aware that having a lawsuit pending often raises the suspicion that people may try to exaggerate their difficulties. This means that your impairments resulting from the head injury must be believable. Major exaggerations, such as not being able to do anything, remembering absolutely nothing, or failing to respond are easy to detect. *At least one of the tests you will be given is designed to catch you faking, because it's easier than it looks. It catches fakers who overdo their impairments.*

In summary, your task is to perform on the tests like someone with a traumatic brain injury. You must perform impaired on the tests of cognitive ability, but not be caught by the tests designed to catch fakers who give extremely low effort.

You have the opportunity to prepare. Before you come to the lab to complete your evaluation, we strongly encourage you to use any resources you can find to help you achieve this goal. Please review the common symptoms of traumatic brain injury that we will be including in the email with theses instructions. We also strongly recommend seeking information from additional sources prior to your testing session (e.g., the internet, the library, friends/family, etc.). Key terms that may help your search and preparation include: malingering, performance validity testing, and neuropsychological evaluations for traumatic brain injury.

If you are able to appear brain injured and avoid being caught as 'a faker,' you will be awarded an additional $20 on the day of participation, and you will be entered into a raffle to win an additional $300.

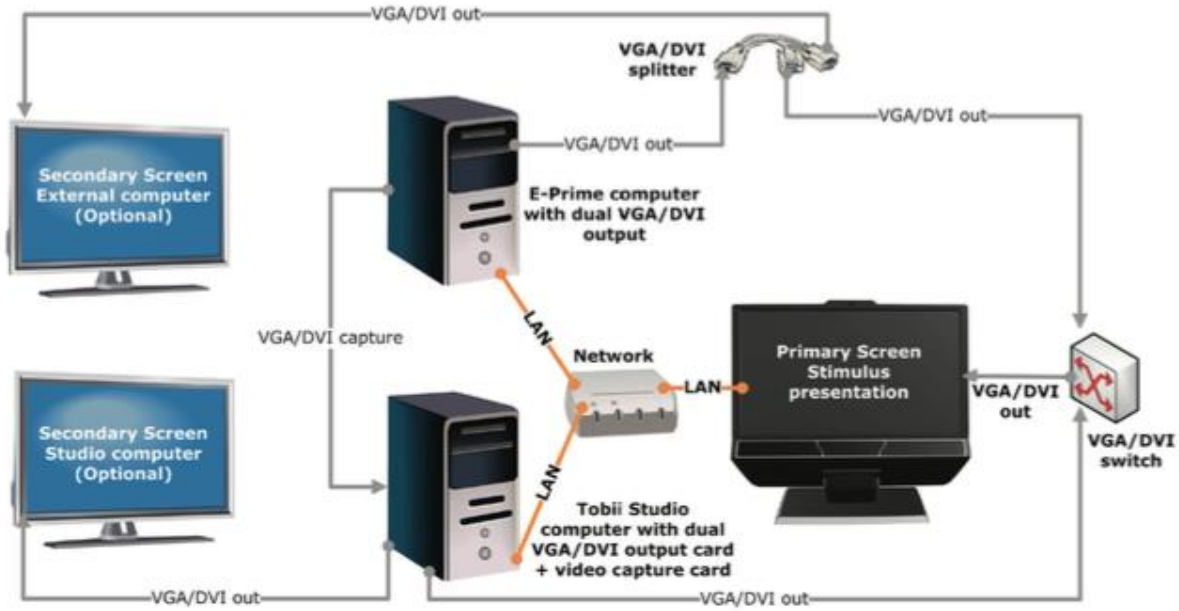
APPENDIX B (FIGURES)

Figure 1. Integrated screen setup example.
Taken from Tobii TX-300 Eye Tracker User Manual, Revision 1 (Tobii Technologies, 2011).

Figure 2. Two-computer set up for Tobii TX-300 integration with Tobii Studio and E-Prime Extensions for Tobii.
Taken from Tobii TX-300 Eye Tracker User Manual, Revision 1 (Tobii Technologies, 2011).

Two-computers setup

In a two-computer setup you will use Tobii Studio to record the gaze data and E-Prime® to present the stimulus. The E-Prime® software and Tobii Studio software are run on two different computers and communicate with the Tobii Eye Tracker Server as well as between each other over a LAN network (TCP/IP protocol).



The image from the Tobii Studio computer has to displayed on the Tobii Eye Tracker's display during calibration (to display the calibration points), whereas the E-Prime® computer should be shown on the Tobii Eye Tracker's display during the actual test. Therefore, both computers have to be connected to the Tobii Eye Tracker's display. Two secondary screens can be added to the setup to enable the eye tracker operator to monitor the progress of the trial during the stimulus presentation and recording.

The image or video presented by E-Prime® on the eye tracker's display also has to be recorded by Tobii Studio and synchronized with the gaze data. The image or video recording is done via a video capture card installed on the Tobii Studio computer. As a result, the display output of the E-Prime® computer needs to be split in two connections: one to the eye tracker's display and another to the video capture card on the computer running Tobii Sudio.

APPENDIX C (TABLES)

Table 1. *Spearman Rho Correlations for Warrington RMT-Words Accuracy and Eye Tracking (ET) Indices as a function of Data Quality (Tobii Validity Index Cutoffs)*

	No Validity Restriction	**Validity > 50%**	**Validity > 60%**	**Validity > 70%**	**Validity > 80%**
Cases lost (%)	0%	6%	9%	13%	29%
Cases Retained (%)	100%	94%	91%	87%	71%
N	174	163	157	150	123
Transitions-Stim	-.16*	-.12	-.12	-.12	-.06
Transitions-FC	-.54**	-.53**	-.53**	-.54**	-.56**
Fixation Count-Stim	.01	-.02	-.02	.00	.04
Fixation Count-FC	-.48**	-.58**	-.60**	-.63**	-.68**
BackgroundFix-Stim[1]	.02	.04	.03	.03	.06
BackgroundFix-FC[1]	-.35**	-.36**	-.36**	-.34**	-.39**
Correct Duration[1]	-.34**	-.38**	-.40**	-.44**	-.52**
Incorrect Duration[1]	-.48**	-.54**	-.56**	-.59**	-.68**
Correct/Incorrect Ratio[1]	.42**	.43**	.44**	.44**	.50**

1. Duration indexes (ms).
*$p < .05$, **$p < .01$

Table 2. Group Differences in Frequency of Excluded Cases Across Tobii Validity Cut Scores.

				% Cases Excluded		
	$X^2(2)^1$	*p*	*phi*	HC	TBI	SIM
Tobii Validity Cut Scores:						
Validity ≥ 50%	5.39	0.07	0.18	3%	13%	2%
Validity ≥ 60%	9.23	0.01	0.23	6%	20%	2%
Validity ≥ 70%	11.88	0.003	0.26	9%	27%	4%
Validity ≥ 80%	11.74	0.003	0.26	20%	46%	21%

Table 3. *Spearman Rho Correlations for Warrington RMT-Words Accuracy and Eye Tracking (ET) Indices (N = 174).*

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. RMT-C	1.00												
2. Transitions-Stim	-.16*	1.00											
3. Transitions-FC	-.54**	.41**	1.00										
4. Fixation Count-Stim	.01	.55**	.17*	1.00									
5. Fixation Count-FC	-.48**	-.07	.34**	.11	1.00								
6. BackgroundFix-Stim[1]	.02	.32**	.02	.23**	-.09	1.00							
7. BackgroundFix-FC[1]	-.35**	.25**	.70**	.18*	.21**	.04	1.00						
8. Correct Duration[1]	-.34**	-.21**	.16*	-.12	.78**	-.19**	-.02	1.00					
9. Incorrect Duration[1]	-.48**	-.15*	.21**	-.08	.83**	-.16*	.05	.92**	1.00				
10. Correct/Incorrect Ratio[1]	.42**	-.04	-.21**	-.05	-.31**	-.02	-.12	-.07	-.40**	1.00			
11. Age	.06	-.10	-.16*	-.14*	-.16*	-.13*	-.13*	.05	-.03	.26**	1.00		
12. Education	.03	-.13*	-.05	.13*	.11	-.08	-.12	.19**	.18	-.07	.11	1.00	
13. WTAR SS	.06	-.09	.02	.11	.09	.00	.00	.11	.09	-.08	-.06	.64**	1.00

Note. RMT-C= Recognition Memory Test Correct; Stim = Stimulus trials; FC = Forced choice trials; BackgroundFix = time spent fixating on background; Correct Duration = time spent looking at the correct forced choice item; Incorrect Duration = time spent looking at the incorrect forced choice item; Correct/Incorrect Ratio = ratio of time spent looking at the correct to incorrect forced choice item; WTAR SS = Wechsler Test of Adult Reading Standard Score.

1. Duration indexes (ms).

*$p < .05$, **$p < .01$

Table 4a. *Kruskal-Wallis Group Comparisons: RMT-Words Performance for HC (n = 67), SIM-IP (n = 47) and TBI (n = 49) Groups.*

Variable	*HC* *Mean Rank*	*SIM* *Mean Rank*	*TBI* *Mean Rank*	*M*	*SD*	*Range*	*Kruskal-Wallis* $X^2(2)$	*p*	Mann-Whitney Contrasts[3]
RMT-C	112.5	42.3	78.3	44.1	6.5	16.0 – 50.0	62.11	< .001	SIM < TBI < HC
Transitions-Stim[1]	77.4	83.0	87.3	6.2	11.9	0.0 – 77.8	1.28	.527	--
Transitions-FC	52.8	112.6	92.6	3.4	4.8	0.3 – 42.0	47.86	< .001	SIM > TBI > HC
Fixation Count-Stim	81.4	82.4	82.5	2.2	0.7	0.7 – 3.8	0.02	.990	--
Fixation Count-FC	60.2	117.0	78.3	3.1	1.2	1.2 – 7.4	40.54	< .001	SIM > TBI > HC
Background Fixation-Stim[2]	88.4	75.5	79.5	290	666	0 – 2964	2.26	.323	--
Background Fixation-FC[2]	52.1	98.6	106.9	120	331	0 – 3448	46.42	< .001	SIM = TBI > HC
Correct Duration[2]	66.8	107.6	78.3	801	438	77 – 3258	21.05	< .001	SIM > TBI = HC
Incorrect Duration[2]	62.6	116.5	75.4	691	424	85 – 3300	37.4	< .001	SIM > TBI = HC
Correct/Incorrect Ratio[2]	94.3	50.5	95.4	1.2	0.2	0.7 – 1.9	29.53	< .001	SIM < TBI = HC

Note. RMT-C = Recognition Memory Test Correct; Stim = Stimulus Items; FC = Forced Choice Items; Background Fixation = time spent fixating on background; Correct Duration = time spent looking at the correct forced choice item; Incorrect Duration = time spent looking at the incorrect forced choice item; Correct/Incorrect Ratio = ratio of time spent looking at the correct to incorrect forced choice item.

1. All oculomotor variables reflect events per trial average.
2. Duration indexes (ms)
3. Mann-Whitney post hoc tests, $p < .05$ criterion.

63

Table 4b. *Mann-Whitney Group Comparisons: RMT-Words Performance for HC (n = 67), SIM-IP (n = 47) and TBI (n = 49) Groups.*

	HC-SIM				SIM-TBI				HC-TBI			
Variable	*U*	*Z*	*p*	*d*	*U*	*Z*	*p*	*d*	*U*	*Z*	*p*	*d*
RMT-C	283.5	-7.47	< .001	1.96	577.5	-4.22	< .001	0.95	887.5	-4.25	< .001	0.86
Transitions-Stim[1]	1458.5	-0.67	.504	0.13	1082.5	-0.51	.613	0.10	1449.5	-1.07	.283	0.20
Transitions-FC	473.0	-6.34	< .001	1.48	817.0	-2.45	.014	0.52	785.5	-4.79	< .001	0.99
Fixation Count-Stim	1551.5	-0.13	.895	0.02	1149.0	-0.02	.985	0.00	1621.5	-0.11	.911	0.02
Fixation Count-FC	531.0	-6.01	< .001	1.36	549.0	-4.42	< .001	1.10	1221.5	-2.35	.019	0.45
Background Fixation-Stim[2]	1330.0	-1.41	.159	0.27	1090.0	-0.45	.652	0.09	1457.5	-1.03	.304	0.19
Background Fixation-FC[2]	661.0	-5.26	< .001	0.29	1020.0	-0.96	.335	0.20	552.0	-6.10	< .001	1.37
Correct Duration[2]	790.0	-4.52	< .001	0.93	735.0	-3.05	.002	0.66	1406.0	-1.32	.188	0.25
Incorrect Duration[2]	567.0	-5.80	< .001	1.29	537.0	-4.50	< .001	1.04	1349.0	-1.64	.102	0.31
Correct/Incorrect Ratio[2]	704.0	-5.01	< .001	1.06	539.0	-4.49	< .001	1.03	1598.0	-0.24	.808	0.05

Note. RMT-C= Recognition Memory Test Correct; Stim = Stimulus Items; FC = Forced Choice Items; Background Fixation = time spent fixating on background; Correct Duration = time spent looking at the correct forced choice item; Incorrect Duration = time spent looking at the incorrect forced choice item; Correct/Incorrect Ratio = ratio of time spent looking at the correct to incorrect forced choice item. Cohen's d estimated from Mann-Whitney Z.

1. All oculomotor events reflect per trial averages.
2. Duration indexes (ms).

Table 5a. *Classification Statistics: RMT-Words Performance for Single-variable Models Predicting Simulator (SIM-IP, n = 47) and Full Effort Healthy Comparison (HC, n = 67) Group Membership*.

	Hit Rate	Sn	Sp	PPP BR 40%	NPP BR 40%	PPP BR 10%	NPP BR 10%	R^2	AUC	AUC 95% CI
One-Variable Models:										
RMT-C	.83	.70	.93	.86	.82	.50	.97	.69	.91	[.85, .97]
Transitions-Stim	.59	.00	1.00	--	.60	--	.90	.00	.53	[.43, .64]
Transitions-FC	.75	.55	.90	.78	.75	.35	.95	.43	.86	[.79, .93]
Fixation Count-Stim	.59	.00	1.00	--	.60	--	.90	.00	.51	[.40, .61]
Fixation Count-FC	.82	.68	.91	.84	.81	.47	.96	.56	.83	[.75, .92]
Background Fixation-Stim	.59	.00	1.00	--	.60	--	.90	.06	.42	[.31, .52]
Background Fixation-FC	.68	.38	.90	.71	.69	.27	.93	.35	.80	[.72, .88]
Correct Duration	.74	.51	.90	.77	.73	.35	.94	.31	.75	[.65, .85]
Incorrect Duration	.80	.66	.90	.81	.79	.42	.96	.46	.82	[.73, .91]
Correct/Incorrect Ratio	.71	.55	.82	.68	.74	.25	.94	.34	.78	[.69, .86]
Two-Variable Models:										
RMT-C + Transitions-FC	.86	.77	.93	.88	.86	.56	.97	.70	.92	[.87, .98]
RMT-C + Fixation Count-FC	.88	.79	.94	.90	.86	.60	.98	.69	.92	[.86, .98]
RMT-C + Background Fixation-Stim	.84	.70	.94	.89	.82	.57	.97	.67	.92	[.87, .98]
RMT-C + Background Fixation-FC	.89	.79	.96	.92	.87	.69	.98	.72	.93	[.88, .98]
RMT-C + Correct Duration	.85	.72	.94	.89	.83	.57	.97	.67	.92	[.86, .97]
RMT-C + Incorrect Duration	.88	.79	.94	.90	.86	.60	.98	.68	.92	[.87, .97]
RMT-C + Correct/Incorrect Ratio	.88	.79	.94	.90	.86	.60	.98	.67	.92	[.86, .98]

Note. Sn = Sensitivity (detection of simulated TBI), Sp = Specificity (identification as HC); 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; RMT = Recognition Memory Test; Stim = Stimulus items; FC = Forced-choice items; Background Fixation = time spent fixating on background; Correct Duration = time spent looking at correct forced-choice item; Incorrect Duration = time spent looking at incorrect forced-choice item; Correct/Incorrect Ratio = ratio of time spent looking at the correct to incorrect forced-choice item.

Table 5b. *Logistic Regressions: RMT-Words Predicting Simulator (SIM-IP, n = 47) and Full Effort Healthy Comparison (HC, n = 67) Group Membership.*

	Df	X^2	*p*	*Odds Ratio*	*Predictors*[1] *p*
One-Variable Models:					
RMT-C	1	75.82	< .001	0.59	
Transitions-Stim	1	0.10	.748	0.99	
Transitions-FC	1	43.15	< .001	2.79	
Fixation Count-Stim	1	0.03	.860	1.05	
Fixation Count-FC	1	51.53	< .001	5.59	
Background Fixation-Stim	1	4.98	.026	0.96	
Background Fixation-FC	1	29.90	< .001	1.50	
Correct Duration	1	30.21	< .001	26.36	
Incorrect Duration	1	47.63	< .001	141.41	
Correct/Incorrect Ratio	1	27.92	< .001	0.94	
Two-Variable Models:					
RMT-C + Transitions-FC	2	84.24	< .001	1.74	.016
RMT-C + Fixation Count-FC	2	81.37	< .001	2.33	.031
RMT-C + Background Fixation-Stim	2	77.85	< .001	0.96	.204
RMT-C + Background Fixation-FC	2	87.01	< .001	1.34	.011
RMT-C + Correct Duration	2	78.45	< .001	4.88	.129
RMT-C + Incorrect Duration	2	80.50	< .001	10.49	.054
RMT-C + Correct/Incorrect Ratio	2	78.71	< .001	0.97	.099

Note. RMT = Recognition Memory Test; Stim = Stimulus items; FC = Forced-choice items; Background Fixation = time spent fixating on background; Correct Duration = time spent looking at correct forced-choice item; Incorrect Duration = time spent looking at incorrect forced-choice item; Correct/Incorrect Ratio = ratio of time spent looking at the correct to incorrect forced-choice item.

1. Refers to eyetracking-variable predictors added on the second step of the two-variable model.

Table 6a. *Classification Statistics: RMT-Words Performance for Single and Two-variable Models Predicting Simulator (SIM, n = 47) and Traumatic Brain Injury (TBI, n = 49) Group Membership.*

	Hit Rate	Sn	Sp	PPP BR 40%	NPP BR 40%	PPP BR 10%	NPP BR 10%	R^2	AUC	AUC 95% CI
One-Variable Models:										
RMT-C	.69	.60	.78	.64	.75	.24	.94	.27	.75	[.65, .85]
Transitions-Stim	.56	.72	.41	.45	.69	.12	.92	.03	.46	[.34, .58]
Transitions-FC	.52	.26	.78	.43	.62	.10	.91	.01	.65	[.54, .76]
Fixation Count-Stim	.51	1.00	.00	.40	--	.10	--	.00	.50	[.38, .62]
Fixation Count-FC	.71	.60	.82	.70	.76	.27	.95	.33	.77	[.67, .87]
Background Fixation-Stim	.52	.64	.41	.42	.63	.11	.92	.01	.46	[.35, .58]
Background Fixation-FC	.52	.64	.40	.42	.62	.10	.92	.00	.44	[.32, .56]
Correct Duration	.64	.53	.74	.57	.70	.19	.93	.17	.68	[.57, .79]
Incorrect Duration	.72	.64	.80	.68	.77	.26	.96	.33	.77	[.67, .87]
Correct/Incorrect Ratio	.69	.70	.67	.59	.76	.19	.95	.27	.77	[.67, .86]
Two-Variable Models:										
RMT-C + Fixation Count-FC	.70	.66	.74	.63	.76	.21	.96	.38	.80	[.71, .89]
RMT-C + Correct Duration	.67	.60	.74	.61	.72	.21	.94	.32	.78	[.69, .87]
RMT-C + Incorrect Duration	.71	.64	.78	.66	.76	.24	.96	.40	.81	[.72, .89]
RMT-C + Correct/Incorrect Ratio	.70	.68	.71	.60	.77	.22	.95	.38	.80	[.71, .88]

Note. Sn = Sensitivity (detection of simulated TBI), Sp = Specificity (identification as HC); 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; RMT = Recognition Memory Test; Stim = Stimulus items; FC = Forced Choice Items; Background Fixation = time spent fixating on the background; Correct Duration = time spent looking at the correct forced choice item; Incorrect Duration = time spent looking at the incorrect forced choice item; Correct/Incorrect Ratio = ratio of time spent looking at the correct to incorrect forced choice item.

Table 6b. *Logistic Regressions Statistics: RMT-Words Predicting Traumatic Brain Injury (TBI) and Simulator (SIM) Group Membership.*

	Df	X^2	*p*	*Odds Ratio*	*Predictors*[1] *p*
One-Variable Models:					
RMT-C	1	21.74	< .001	0.85	
Transitions-Stim	1	2.41	.136	0.95	
Transitions-FC	1	0.97	.330	1.07	
Fixation Count-Stim	1	0.00	.973	1.01	
Fixation Count-FC	1	27.05	< .001	2.97	
Background Fixation-Stim	1	0.47	.496	0.98	
Background Fixation-FC	1	0.08	.778	0.99	
Correct Duration	1	13.35	< .001	5.95	
Incorrect Duration	1	27.62	< .001	26.54	
Correct/Incorrect Ratio	1	21.69	< .001	0.95	
Two-Variable Models:					
RMT-C + Fixation Count-FC	2	32.10	< .001	2.27	.004
RMT-C + Correct Duration	2	26.47	< .001	3.32	.041
RMT-C + Incorrect Duration	2	33.78	< .001	11.49	.003
RMT-C + Correct/Incorrect Ratio	2	31.85	< .001	0.96	.004

Note. RMT = Recognition Memory Test; Stim = Stimulus items; FC = Forced-choice items; Background Fixation = time spent fixating on background; Correct Duration = time spent looking at correct forced-choice item; Incorrect Duration = time spent looking at incorrect forced-choice item; Correct/Incorrect Ratio = ratio of time spent looking at the correct to incorrect forced-choice item.
1. Refers to eyetracking-variable predictors added on the second step of the two-variable model.

Table 7a. *Group Differences for SIM-IP (n = 49) and SIM-SC (n = 46) Simulator Conditions in Demographics, Neuropsychological Test Performance (T scores), and Rates of Impairment.*

	SIM-IP *M* (SD)	SIM-SC *M* (SD)	Cohen's d	*phi*
Age (years)	34.0 (16.4)	43.6 (16.2)	.58**	
Education (years)	14.9 (2.2)	14.5 (2.0)	.14	
Estimated FSIQ (WTAR)	106.3 (13.3)	101.7 (15.2)	.32	
Trails A	28.8 (17.6)	34.2 (14.8)	.33	
(% impaired t < 40)	79.6	60.9		.21*
(% impaired t < 35)	67.3	52.2		.16
Trails B	33.4 (14.0)	42.3 (13.4)	.64**	
(% impaired t < 40)	67.3	43.5		.24**
(% impaired t < 35)	55.1	30.4		.25**
SDMT	25.0 (15.0)	26.6 (14.4)	.11	
(% impaired t < 40)	87.8	84.8		.04
(% impaired t < 35)	69.4	73.9		-.05
Digit Span FW	36.8 (10.9)	36.2 (10.1)	.06	
(% impaired t < 40)	61.2	65.2		-.04
(% impaired t < 35)	42.9	52.2		-.09
Digit Span BW	39.5 (9.1)	39.9 (8.9)	.04	
(% impaired t < 40)	55.1	67.4		-.13
(% impaired t < 35)	22.4	19.6		.04
Digit Span SQ	39.4 (10.2)	38.6 (9.0)	.08	
(% impaired t < 40)	65.3	71.7		-.07
(% impaired t < 35)	30.6	28.3		.03
CVLT Trials 1-5	38.7 (14.3)	35.6 (11.2)	.24	
(% impaired t < 40)	55.1	63.6		-.09
(% impaired t < 35)	40.8	50.0		-.09
CVLT SDFR	34.5 (14.3)	35.7 (10.9)	.09	
(% impaired t < 40)	61.2	70.5		-.10
(% impaired t < 35)	55.1	59.1		-.04
CVLT LDFR	31.4 (13.8)	33.3 (13.1)	.14	
(% impaired t < 40)	77.6	70.5		.08
(% impaired t < 35)	65.3	61.4		.04

Note: SIM-IP = financial incentive + time to prepare condition; SIM-SC = single coaching session condition; SDMT = Symbol Digit Modalities Test; FW = Forward; BW = Backward; SQ = Sequencing; CVLT = California Verbal Learning Test-II; SDFR = Short Delay Free Recall; LDFR = Long Delay Free Recall

*$p < .05$, **$p < .01$, one-tailed test

Table 7b. *Group Differences for SIM-IP and SIM-SC Simulator Conditions in Performance Validity Test Performance, Rates of Failure, and Rates of Successful Simulation.*

	SIM-IP *M* (SD)	SIM-SC *M* (SD)	Cohen's d	*phi*
Performance Validity Tests				
TOMM Trial 1	36.0 (8.2)	32.6 (9.3)	.38*	
(% Failed)	67.3	83.7		.19*
TOMM Trial 2	37.8 (10.5)	35.2 (11.1)	.24	
(% Failed)	63.3	76.7		.15
RMT-C	37.9 (7.9)	35.7 (13.4)	.24	
(% Failed)	18.4	31.1		.15
RDS	6.8 (2.3)	6.8 (2.3)	.00	
(% Failed)	40.8	47.8		.07
CVLT-FC	13.6 (2.6)	13.3 (3.5)	.10	
(% Failed)	57.1	47.7		-.09
Successful Simulation				
% Successful with $t < 40$	40.8	39.1		.02
% Successful with $t < 35$	34.7	32.6		.02

Note: SIM-IP = financial incentive + time to prepare condition; SIM-SC = single coaching session condition; TOMM = Test of Memory Malingering; RMT = Recognition Memory Test; RDS = Reliable Digit Span; CVLT = California Verbal Learning Test-II; FC = Forced Choice
*$p < .05$, **$p < .01$, one-tailed test

REFERENCES

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: a reprocessing effect in face perception. *J Exp Psychol Learn Mem Cogn, 25*(4), 997-1010.

Althoff, R. R., Cohen, N. J., McConkie, G., Wasserman, S., Maciukenas, M., Azen, R., & Romine, L. (1998). Eye-movement based memory assessment. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current oculomotor research: Physiological and Psychological Aspects* (pp. 293–302). New York: Kluwer Academic/Plenum Press.

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

Antoniades, C. A., & Kennard, C. (2015). Ocular motor abnormalities in neurodegenerative disorders. *Eye (Lond), 29*(2), 200-207. doi:10.1038/eye.2014.276

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. *Clin Neuropsychol, 15*(4), 516-520. doi:10.1076/clin.15.4.516.1884

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. *Int J Neurosci, 112*(12), 1479-1487.

Backhaus, S. L., Fichtenberg, N. L., & Hanks, R. A. (2004). Detection of sub-optimal performance using a floor effect strategy in patients with traumatic brain injury. *Clinical Neuropsychologist, 18*(4), 591-603. doi:10.1080/13854040490888558

Bashem, J. R., Rapport, L. J., Kanser, R. J., Billings, N., Vermilion, B., Krohner, S., . . . Siple, P. (2017). Performance validity assessment of bona fide and malingered traumatic brain

injury using novel eye-tracking systems. *Journal of the International Neuropsychological Society, 23*(S1), 32.

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? *Arch Clin Neuropsychol, 21*(1), 121-126. doi:10.1016/j.acn.2005.06.010

Bender, S. D., & Rogers, R. (2004). Detection of neurocognitive feigning: development of a multi-strategy assessment. *Arch Clin Neuropsychol, 19*(1), 49-60.

Bernard, L. C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *J Clin Exp Neuropsychol, 12*(5), 715-728. doi:10.1080/01688639008401014

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: a critical review. *Clin Neuropsychol, 15*(1), 19-45. doi:10.1076/clin.15.1.19.1907

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. *Am J Psychiatry, 153*(1), 7-10.

Bittencourt, J., Velasques, B., Teixeira, S., Basile, L. F., Salles, J. I., Nardi, A. E., . . . Ribeiro, P. (2013). Saccadic eye movement applications for psychiatric disorders. *Neuropsychiatr Dis Treat, 9*, 1393-1409. doi:10.2147/NDT.S45931

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. *J Clin Exp Neuropsychol, 24*(2), 154-167. doi:10.1076/jcen.24.2.154.1000

Brennan, A. M., Meyer, S., David, E., Pella, R., Hill, B. D., & Gouvier, W. D. (2009). The vulnerability to coaching across measures of effort. *Clin Neuropsychol, 23*(2), 314-328. doi:10.1080/13854040802054151

Cassar, J. R., Hales, E. S., Longhurst, J. G., & Weiss, G. S. (1996). Can disability benefits make children sicker? *J Am Acad Child Adolesc Psychiatry, 35*(6), 700-701. doi:10.1097/00004583-199606000-00005

Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., . . . Ord, J. S. (2015). Official position of the American Academy of Clinical Neuropsychology Social Security Administration Policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist, 29*(6), 723-740. doi:10.1080/13854046.2015.1099738

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. *Arch Phys Med Rehabil, 92*(4), 519-530. doi:10.1016/j.apmr.2010.11.015

Coleman, R. D., Rapport, L. J., Millis, S. R., Ricker, J. H., & Farchione, T. J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *J Clin Exp Neuropsychol, 20*(2), 201-210. doi:10.1076/jcen.20.2.201.1164

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? *Arch Clin Neuropsychol, 20*(2), 191-198. doi:10.1016/j.acn.2004.06.002

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychol, 9*(2), 81-90. doi:10.1076/chin.9.2.81.14505

Cook, A. E., Hacker, D. J., Webb, A. K., Osher, D., Kristjansson, S. D., Woltz, D. J., & Kircher, J. C. (2012). Lyin' eyes: ocular-motor measures of reading reveal deception. *J Exp Psychol Appl, 18*(3), 301-313. doi:10.1037/a0028307

Coronado, V. G., McGuire, L. C., Sarmiento, K., Bell, J., Lionbarger, M. R., Jones, C. D., . . . Xu, L. (2012). Trends in Traumatic Brain Injury in the U.S. and the public health response: 1995-2009. *J Safety Res, 43*(4), 299-307. doi:10.1016/j.jsr.2012.08.011

Corso, P., Finkelstein, E., Miller, T., Fiebelkorn, I., & Zaloshnja, E. (2006). Incidence and lifetime costs of injuries in the United States. *Inj Prev, 12*(4), 212-218. doi:10.1136/ip.2005.010983

Corso, P., Finkelstein, E., Miller, T., Fiebelkorn, I., & Zaloshnja, E. (2015). Incidence and lifetime costs of injuries in the United States. *Inj Prev, 21*(6), 434-440. doi:10.1136/ip.2005.010983rep

Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *J Clin Exp Neuropsychol, 39*(3), 257-264. doi:10.1080/13803395.2016.1223278

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (2000). The California Verbal Learning Test (2nd ed.): Psychological Assessment Resources, Inc.

Denning, J. H. (2012). The efficiency and accuracy of the Test of Memory Malingering trial 1, errors on the first 10 items of the test of memory malingering, and five embedded measures in predicting invalid test performance. *Arch Clin Neuropsychol, 27*(4), 417-432. doi:10.1093/arclin/acs044

Dikmen, S. S., Corrigan, J. D., Levin, H. S., Machamer, J., Stiers, W., & Weisskopf, M. G. (2009). Cognitive outcome following traumatic brain injury. *J Head Trauma Rehabil, 24*(6), 430-438. doi:10.1097/HTR.0b013e3181c133e9

Draper, K., & Ponsford, J. (2008). Cognitive functioning ten years following traumatic brain injury and rehabilitation. *Neuropsychology, 22*(5), 618-625. doi:10.1037/0894-4105.22.5.618

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behav Res Methods Instrum Comput, 34*(4), 455-470.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science, 10*(3), 263-266. doi:10.1111/1467-9280.00147

Elhai, J. D., Butcher, J. J., Reeves, A. N., Baugher, S. N., Gray, M. J., Jacobs, G. A., . . . Arbisi, P. (2007). Varying cautionary instructions, monetary incentives, and comorbid diagnostic training in malingered psychopathology research. *J Pers Assess, 88*(3), 328-337. doi:10.1080/00223890701332136

Erdodi, L. A., Tyson, B. T., Shahein, A. G., Lichtenstein, J. D., Abeare, C. A., Pelletier, C. L., . . . Roth, R. M. (2017). The power of timing: Adding a time-to-completion cutoff to the Word Choice Test and Recognition Memory Test improves classification accuracy. *J Clin Exp Neuropsychol, 39*(4), 369-383. doi:10.1080/13803395.2016.1230181

Essig, S. M., Mittenberg, W., Petersen, R. S., Strauman, S., & Cooper, J. T. (2001). Practices in forensic neuropsychology: perspectives of neuropsychologists and trial attorneys. *Arch Clin Neuropsychol, 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*. Retrieved from Atlanta, GA:

Faust, D. (1995). The detection of deception. *Neurol Clin, 13*(2), 255-265.

Frederick, R. I., Sarfaty, S. D., Johnston, J. D., & Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology, 8*(1), 118-125.

Goldberg, H., Back-Madruga, C., & Boone, K. B. (2007). The impact of psychiatric disorders on cognitive symptom validity test scores. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 281-309). New York: Guilford Press.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics, 24*(6), 631-645. doi:10.1016/s0169-8141(98)00068-7

Gouvier, W. D., Prestholdt, P. H., & Warner, M. S. (1988). A survey of common misconceptions about head injury and recovery. *Arch Clin Neuropsychol, 3*(4), 331-343.

Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms, US Version 1.1*. Durham, NC: CogniSyst.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M., 3rd. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Inj, 15*(12), 1045-1060. doi:10.1080/02699050110088254

Greiffenstein, M. F., Baker, W.J., & Gola, T. (1994). Validation of malingered amnesia measures in a large clinical sample. *Psychological Assessment, 6*, 218-224.

Griffin, Z. M., & Oppenheimer, D. M. (2006). Speakers gaze at objects while preparing intentionally inaccurate labels for them. *J Exp Psychol Learn Mem Cogn, 32*(4), 943-948. doi:10.1037/0278-7393.32.4.943

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience, 4*. doi:10.3389/fnhum.2010.00166

Hannula, D. E., & Ranganath, C. (2009). The eyes have it: hippocampal activity predicts expression of memory in eye movements. *Neuron, 63*(5), 592-599. doi:10.1016/j.neuron.2009.08.025

Hannula, D. E., Ryan, J. D., Tranel, D., & Cohen, N. J. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *J Cogn Neurosci, 19*(10), 1690-1705. doi:10.1162/jocn.2007.19.10.1690

Heaton, R. K., Smith, H. H., Jr., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *J Consult Clin Psychol, 46*(5), 892-900.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference, P. (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. doi:10.1080/13854040903155063

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

Iverson, G. L., & Franzen, M. D. (1994). The Recognition Memory Test, Digit Span, and Knox Cube Test as markers of malingered memory impairment. *Assessment, 1*(4), 323–334.

Iverson, G. L., & Franzen, M. D. (1998). Detecting malingered memory deficits with the Recognition Memory Test. *Brain Inj, 12*(4), 275-282.

Jacob, R. J., & Karn, J. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyona, R. Radach, & H. Deubel (Eds.), *Mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573-605).

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441-480. doi:10.1016/0010-0285(76)90015-3

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

Kanser, R. J., Rapport, L. J., Bashem, J. R., Billings, N. M., Hanks, R. A., Axelrod, B. N., & Miller, J. B. (2017). Strategies of successful and unsuccessful simulators coached to feign traumatic brain injury. *Clin Neuropsychol, 31*(3), 644-653. doi:10.1080/13854046.2016.1278040

Kanser, R. J., Rapport, L. J., Bashem, J. R., & Hanks, R. A. (2018). Detecting malingering in traumatic brain injury: combining response time with performance validity test accuracy. *Clin Neuropsychol*, 1-18. doi:10.1080/13854046.2018.1440006

Kim, M. S., Boone, K. B., Victor, T., Marion, S. D., Amano, S., Cottingham, M. E., . . . Zeller, M. A. (2010). The Warrington Recognition Memory Test for words as a measure of response bias: total score and response time cutoffs developed on "real world" credible and noncredible subjects. *Arch Clin Neuropsychol, 25*(1), 60-70. doi:10.1093/arclin/acp088

Langlois, J. A., Rutland-Brown, W., & Wald, M. M. (2006). The epidemiology and impact of traumatic brain injury: a brief overview. *J Head Trauma Rehabil, 21*(5), 375-378.

Langlois Orman, J. L., Kraus, J. F., Zaloshnja, E., & Miller, T. (2011). Epidemiology. In J. M. Silver, T. W. McAllister, & S. C. Yudofsky (Eds.), *Textbook of traumatic brain injury* (Second ed., pp. 3-22). Washington, DC: American Psychiatric Publishing, Inc.

Larrabee, G. J. (2005). Mild traumatic brain injury. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (pp. 209–236). New York: Oxford University Press.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *J Int Neuropsychol Soc, 18*(4), 625-630.

Larrabee, G. J. (Ed.) (2011). *Forensic neuropsychology: A scientific approach* (Second ed.). New York: Oxford University Press.

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–371). New York: Oxford.

Larrabee, G. J., Greiffenstein, M.F., Greve, K.W. & Bianchini, K.J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–371). New York: Oxford.

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009a). 40 plus or minus 10, a new magical number: reply to Russell. *Clin Neuropsychol, 23*(5), 841-849. doi:10.1080/13854040902796735

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009b). 40 plus or minus 10, a new magical number: reply to Russell. *The Clinical Neuropsychologist, 23*(5), 841-849. doi:10.1080/13854040902796735

Lau, L., Basso, M. R., Estevis, E., Miller, A., Whiteside, D. M., Combs, D., & Arentsen, T. J. (2017). Detecting coached neuropsychological dysfunction: a simulation experiment regarding mild traumatic brain injury. *Clin Neuropsychol, 31*(8), 1412-1431. doi:10.1080/13854046.2017.1318954

Leighton, A., Weinborn, M., & Maybery, M. (2014). Bridging the gap between neurocognitive processing theory and performance validity assessment among the cognitively impaired: a review and methodological approach. *J Int Neuropsychol Soc, 20*(9), 873-886. doi:10.1017/S135561771400085X

Lippa, S. M. (2018). Performance validity testing in neuropsychology: a clinical guide, critical review, and update on a rapidly evolving literature. *Clin Neuropsychol, 32*(3), 391-421. doi:10.1080/13854046.2017.1406146

Lu, P. H., & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: malingering by proxy? *Clin Neuropsychol, 16*(1), 90-96. doi:10.1076/clin.16.1.90.8328

Lupu, T., Elbaum, T., Wagner, M., & Braw, Y. (2018). Enhanced detection of feigned cognitive impairment using per item response time measurements in the Word Memory Test. *Appl Neuropsychol Adult, 25*(6), 532-542. doi:10.1080/23279095.2017.1341410

Mahoney, E. J., Kapur, N., Osmon, D. C., & Hannula, D. E. (2018). Eye Tracking as a Tool for the Detection of Simulated Memory Impairment. *Journal of Applied Research in Memory and Cognition, 7*(3), 441-453. doi:10.1016/j.jarmac.2018.05.004

Mani, R., Asper, L., & Khuu, S. K. (2018). Deficits in saccades and smooth-pursuit eye movements in adults with traumatic brain injury: a systematic review and meta-analysis. *Brain Inj, 32*(11), 1315-1336. doi:10.1080/02699052.2018.1483030

Manns, J. R., Stark, C. E., & Squire, L. R. (2000). The visual paired-comparison task as a measure of declarative memory. *Proc Natl Acad Sci U S A, 97*(22), 12375-12379. doi:10.1073/pnas.220398097

Martin, R. C., Bolter, J. F., Todd, M. E., Gouvier, W. D., & Niccolls, R. (1993). Effects of sophistication and motivation on the detection of malingered memory performance using a computerized forced-choice task. *J Clin Exp Neuropsychol, 15*(6), 867-880. doi:10.1080/01688639308402604

Meyers, J., Reinsch-Boothby, L., Miller, R., Rohling, M., & Axelrod, B. (2011). Does the source of a forensic referral affect neuropsychological test performance on a standardized battery of tests? *Clin Neuropsychol, 25*(3), 477-487. doi:10.1080/13854046.2011.554442

Millen, A. E., Hope, L., Hillstrom, A. P., & Vrij, A. (2017). Tracking the truth: the effect of face familiarity on eye fixations during deception. *Q J Exp Psychol (Hove), 70*(5), 930-943. doi:10.1080/17470218.2016.1172093

Millis, S. R. (1992). The recognition memory test in the detection of malingered and exaggerated memory deficits. *The Clinical Neuropsychologist, 6*(4), 406–414.

Millis, S. R. (1994). Assessment of motivation and memory with the Recognition Memory Test after financially compensable mild head injury. *J Clin Psychol, 50*(4), 601-605.

Millis, S. R. (2008). Assessment of incomplete effort and malingering in the neuropsychological evaluation. In J. E. Morgan & J. H. 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 J. E. Morgan

& J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 21-37). New York: Psychology Press.

Millis, S. R., & Putnam, S. H. (1994). The Recognition Memory Test in the assessment of memory impairment after financially compensable mild head injury: A replication. *Perceptual and Motor Skills, 79*(1), 384-386.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *J Clin Exp Neuropsychol, 24*(8), 1094-1102. doi:10.1076/jcen.24.8.1094.8379

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *Clin Neuropsychol, 17*(2), 263-272. doi:10.1076/clin.17.2.263.16511

Neudecker, J. J., & Skeel, R. L. (2009). Development of a novel malingering detection method involving multiple detection strategies. *Arch Clin Neuropsychol, 24*(1), 59-70. doi:10.1093/arclin/acp008

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: a review on eye movements in decision making. *Acta Psychol (Amst), 144*(1), 190-206. doi:10.1016/j.actpsy.2013.06.003

Page, M. P. A. (2006). What can't functional neuroimaging tell the cognitive psychologist? *Cortex, 42*(3), 428-443. doi:10.1016/s0010-9452(08)70375-7

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Percept Mot Skills, 57*(2), 367-372. doi:10.2466/pms.1983.57.2.367

Pankratz, L., Fausti, A., & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *J Consult Clin Psychol, 43*(3), 421-422.

Proudfoot, J. G., Jenkins, J. L., Burgoon, J. K., & Nunamaker, J. F. (2016). More Than Meets the Eye: How Oculometric Behaviors Evolve Over the Course of Automated Deception Detection Interactions. *Journal of Management Information Systems, 33*(2), 332-360. doi:10.1080/07421222.2016.1205929

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol Bull, 124*(3), 372-422.

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. doi:10.1037/1040-3590.10.1.10

Reicker, L. I. (2008). The ability of reaction time tests to detect simulation: an investigation of contextual effects and criterion scores. *Arch Clin Neuropsychol, 23*(4), 419-431. doi:10.1016/j.acn.2008.02.003

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation*. Tucson, AZ: Neuropsychology Press.

Rogers, R. (1990). Models of feigned mental illness. *Professional Psychology-Research and Practice, 21*(3), 182-188.

Rogers, R. (1997). Researching dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (Second ed., pp. 398-426). New York: Guilford Press.

Rogers, R. (2008). An introduction to response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (Third ed., pp. 3-13). New York: Guilford Press.

Rose, F. E., Hall, S., Szalda-Petree, A. D., & Bach, P. J. (1998). A comparison of four tests of malingering and the effects of coaching. *Arch Clin Neuropsychol, 13*(4), 349-363.

Rose, F. E., Hall, S., & Szaldapetree, A. D. (1995). Portland Digit Recognition Test-Computerized - Measuring response latency improves the detection of malingering. *Clinical Neuropsychologist, 9*(2), 124-134. doi:10.1080/13854049508401594

Ross, S. R., Putnam, S. H., & Adams, K. M. (2006). Psychological disturbance, incomplete effort, and compensation-seeking status as predictors of neuropsychological test performance in head injury. *J Clin Exp Neuropsychol, 28*(1), 111-125. doi:10.1080/13803390590929270

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.* (pp. 315–338). Lisse, Netherlands: Swets and Zeitlinger.

Ruiz, M. A., Drake, E. B., Glass, A., Marcotte, D., & van Gorp, W. G. (2002). Trying to beat the system: Misuse of the Internet to assist in avoiding the detection of psychological symptom dissimulation. *Professional Psychology-Research and Practice, 33*(3), 294-299. doi:10.1037//0735-7028.33.3.294

Russeler, J., Brett, A., Klaue, U., Sailer, M., & Munte, T. F. (2008). The effect of coaching on the simulated malingering of memory impairment. *BMC Neurol, 8*, 37. doi:10.1186/1471-2377-8-37

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science, 11*(6), 454-461. doi:10.1111/1467-9280.00288

Ryan, J. D., Moses, S. N., & Villate, C. (2009). Impaired relational organization of propositions, but intact transitive inference, in aging: Implications for understanding underlying neural integrity. *Neuropsychologia, 47*(2), 338-353. doi:10.1016/j.neuropsychologia.2008.09.006

Shum, D. H., O'Gorman, J. G., & Alpar, A. (2004). Effects of incentive and preparation time on performance and classification accuracy of standard and malingering-specific memory tests. *Arch Clin Neuropsychol, 19*(6), 817-823. doi:10.1016/j.acn.2003.10.002

Slick, D. J., Hopp, G., Strauss, E., & Thompson, G. (1997). *The Victoria symptom validity test*. Odessa, FL: Psychological Assessment Resources.

Slick, D. J., & Sherman, E. (2012). Differential diagnosis of malingering. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: System validity assessment and malingering* (pp. 57-72). New York: Springer.

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol, 13*(4), 545-561. doi:10.1076/1385-4046(199911)13:04;1-Y;FT545

Smith, A. (1973). *Symbol Digits Modality Test*. Los Angeles: Western Psychological Services.

Snyder, K. A., Blank, M. P., & Marsolek, C. J. (2008). What form of memory underlies novelty preferences? *Psychon Bull Rev, 15*(2), 315-321.

Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: a meta-analytic update and extension. *Arch Clin Neuropsychol, 26*(8), 774-789. doi:10.1093/arclin/acr066

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. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Arch Clin Neuropsychol, 15*(5), 415-424.

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.

Sweet, J. J., Benson, L. M., Nelson, N. W., & Moberg, P. J. (2015). The American Academy of Clinical Neuropsychology, National Academy of Neuropsychology, and Society for Clinical Neuropsychology (APA Division 40) 2015 TCN Professional Practice and 'Salary Survey': Professional Practices, Beliefs, and Incomes of U.S. Neuropsychologists. *The Clinical Neuropsychologist, 29*(8), 1069-1162. doi:10.1080/13854046.2016.1140228

Swift, T. J., Nicks, R. C., Swift, R. R., Juan, R., & Aguerrevere, L. E. (2017). Detecting simulated memory malingering with eye-tracking technology. *Journal of the International Neuropsychological Society, 23*(S1), 198.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston, MA: Allyn & Bacon/Pearson Education.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *Clin Neuropsychol, 16*(4), 495-505. doi:10.1076/clin.16.4.495.13909

Tardif, H. P., Barry, R. J., Fox, A. M., & Johnstone, S. J. (2000). Detection of feigned recognition memory impairment using the old/new effect of the event-related potential. *Int J Psychophysiol, 36*(1), 1-9.

Taylor, C. A., Bell, J. M., Breiding, M. J., & Xu, L. (2017). Traumatic Brain Injury-Related Emergency Department Visits, Hospitalizations, and Deaths - United States, 2007 and 2013. *MMWR Surveill Summ, 66*(9), 1-16. doi:10.15585/mmwr.ss6609a1

Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Arch Clin Neuropsychol, 19*(3), 455-464. doi:10.1016/S0887-6177(03)00078-7

The Psychological Corporation. (2001). *Wechsler Test of Adult Reading (WTAR)*. San Antonio, TX: The Psychological Corporation.

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

Tombaugh, T. N. (1997). The test of memory malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*(3), 260-268. doi:10.1037/1040-3590.9.3.260

Tombaugh, T. N., & Rees, L. M. (2000). *Manual for the computerized tests of information processing (CTIP)*. Ottawa, ON: Carleton University.

Vagnini, V. L., Berry, D. T., Clark, J. A., & Jiang, Y. (2008). New measures to detect malingered neurocognitive deficit: applying reaction time and event-related potentials. *J Clin Exp Neuropsychol, 30*(7), 766-776. doi:10.1080/13803390701754746

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. *Int J Psychophysiol, 73*(3), 246-256. doi:10.1016/j.ijpsycho.2009.04.002

van Hooft, E. A., & Born, M. P. (2012). Intentional response distortion on personality tests: using eye-tracking to understand response processes when faking. *J Appl Psychol, 97*(2), 301-316. doi:10.1037/a0025711

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the Liars: Toward a Cognitive Lie Detection Approach. *Current Directions in Psychological Science, 20*(1), 28-32. doi:10.1177/0963721410391245

Warrington, E. K. (1984). *Recognition Memory Test*. Lutz, FL: Psychological Assessment Resources.

Wasyliw, O. E., & Golden, C. J. (1985). Neuropsychological evaluation in the assessment of personal injury. *Behavioral Sciences and the Law, 3*(2), 149-164.

Weinborn, M., Woods, S. P., Nulsen, C., & Leighton, A. (2012). The effects of coaching on the verbal and nonverbal medical symptom validity tests. *Clin Neuropsychol, 26*(5), 832-849. doi:10.1080/13854046.2012.686630

Wetter, M. W., & Corrigan, S. K. (1995). Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. *Professional Psychology-Research and Practice, 26*(5), 474-477. doi:10.1037/0735-7028.26.5.474

Willison, J., & Tombaugh, T. N. (2006). Detecting simulation of attention deficits using reaction time tests. *Arch Clin Neuropsychol, 21*(1), 41-52. doi:10.1016/j.acn.2005.07.005

Woods, D. L., Wyma, J. M., Yund, E. W., & Herron, T. J. (2015). The Effects of Repeated Testing, Simulated Malingering, and Traumatic Brain Injury on Visual Choice Reaction Time. *Front Hum Neurosci, 9*, 595. doi:10.3389/fnhum.2015.00595

ABSTRACT

THE EYES NEVER LIE: DETECTING SIMULATED TRAUMATIC BRAIN INJURY WITH EYE-TRACKING

by

ROBERT J KANSER

MAY 2019

Advisor: Dr. Lisa J. Rapport

Major: Psychology (Clinical)

Degree: Doctor of Philosophy

Performance validity test (PVT) inaccuracies can be explained by both test and extra-test (e.g., research design components) factors. Eye-tracking is a promising technology to enhance assessment of performance validity. Prior research has established that ocular behaviors are reliable biomarkers of (un)conscious cognitive processes. Experimental research on deception has shown that ocular behaviors reliably distinguish feigned concealment of information from honest responding. The primary objective of this study was to examine the incremental utility of incorporating eye-tracking into a clinical PVT to distinguish adults with verified TBI from adults coached to feign cognitive impairment. A secondary objective was to determine the effect of financial incentive and preparation time on individuals' ability to simulate cognitive impairment. Participants were adults with moderate to severe TBI (TBI; n = 49), healthy adults coached to simulate TBI in a single session (SIM-SC; n = 46) and those given time to prepare and financial incentive (SIM-IP; n = 49), and healthy adult comparisons providing full effort (HC; n = 67). A computerized version of the Warrington Recognition Memory Test of Words (RMT) was completed in the context of a full neuropsychological battery.

Eye-tracking indices generally did not differ among the groups during presentation of stimulus items but did differ during decision-making (forced-choice) trials. Compared to TBI and HC, SIM-IP had significantly more transitions, fixations, and time spent looking at correct and incorrect response options. Logistic regressions and ROC curve analyses showed that accuracy was the best predictor of SIM-IP vs. HC. For SIM-IP vs. TBI, eye-tracking variables exceeded accuracy in distinguishing the groups. Eye-tracking added incremental predictive value to accuracy in discriminating SIM-IP from HC and TBI.

SIM-IP performed worse than SIM-SC on tasks of processing speed and executive functioning; the two simulation groups performed equivalently on all other cognitive tests. SIM-IP performed better than SIM-SC across forced-choice PVTs with small effects; however, groups did not differ with respect to rates of successful simulation. Overall, the combined effect of preparation time and financial incentive led to a pattern of neuropsychological test performance that was more consistent with patterns observed in TBI. Despite this finding, individuals receiving incentive and preparation time were not more likely to accomplish the challenging goal of successful simulation, which requires performing impaired on tests of cognitive ability while at the same time avoiding PVT detection.

Eye-tracking indicated that persons feigning TBI showed multiple signs of greater cognitive effort than persons with verified TBI and healthy comparisons. Effectiveness of RMT accuracy and eye-tracking depended on the groups compared. In the comparison of greatest interest (SIM vs. TBI) eye-tracking best predicted group status and led to "excellent" discrimination when combined with accuracy. Successful simulation was achieved by 32 – 40% of the simulators, depending on the criterion for "impairment;" remarkably, eye-tracking indices

improved detection of feigning even among this group. Eye-tracking may be an important complement to traditional accuracy scores on PVTs.

AUTOBIOGRAPHICAL STATEMENT

ROBERT J. KANSER

Education

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

08/2013 - Present **Master of Arts**
Wayne State University
Major: Clinical Psychology

Clinical Experience

09/2015 – 09/2016 **Center for Forensic Psychiatry, Ann Arbor, MI**
Supervisors: Judith Shazer, Ph.D., and Jay Witherell, Ph.D

09/2016 – 06/2018 **The Rehabilitation Institute of Michigan, Detroit, MI**
Supervisors: Robin Hanks, Ph.D. ABPP-CN; Tom Gola, Ph.D.; Stephen Vangel, Ph.D.

07/2018 – Present **James A. Haley Veterans' Hospital, Tampa, FL**
Supervisors: Jessica Vassallo, Ph.D., ABPP-CN; Katherine Burns, Ph.D.; Tracy Kretzmer, Ph.D; Marc Silva, Ph.D.

Research Experience

05/2012 – 06/2013 **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.

08/2013 – Present **Research Assistant**
Wayne State University Psychology Department, Detroit, MI
Advisor: Lisa Rapport, Ph.D.

Honors and Awards

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

2017; 2018 The Clinical Neuropsychologist (TCN)/American Academy of Clinical Neuropsychology (AACN) Student Project Competition, runner up and first prize

2018 Steven A. Lewis Memorial Award