2020

DETECTING SIMULATED MEMORY IMPAIRMENT IN COLLEGE STUDENTS WITH THE PROGRESSIVE VISUAL MEMORY TEST (PVMT): VALIDATION OF A NEW TEST OF PERFORMANCE VALIDITY

Brook Elizabeth Clark

DETECTING SIMULATED MEMORY IMPAIRMENT IN COLLEGE STUDENTS WITH THE PROGRESSIVE VISUAL MEMORY TEST (PVMT): VALIDATION OF A NEW TEST OF PERFORMANCE VALIDITY

By

BROOK ELIZABETH CLARK

Master of Arts, University of Montana, Missoula, Montana, 2017
Bachelor of Arts, University of Montana, Missoula, Montana, 2001

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Clinical Psychology, exposure in Neuropsychology

The University of Montana
Missoula, MT

May 2020

Approved by:

Scott Whittenburg,
Graduate School Dean

Stuart Hall, Ph.D., Chair
Department of Psychology

Laura Kirsch, Ph.D.
Department of Psychology

Craig McFarland, Ph.D.
Department of Psychology

Annie Sondag
School of Public and Community Health Sciences

Allen Szalda-Petree
Chair of Department of Psychology

Clark, Brook, Ph.D., May 2020 Clinical Psychology

Detecting Simulated Memory Impairment in College Students with The Progressive Visual Memory Test (PVMT): Validation of a New Test of Performance Validity

Chairperson: Stuart Hall, Ph.D.

The detection of exaggerated or feigned symptoms is a complex issue that neuropsychologists regularly encounter during neuropsychological evaluations. Thus, it is critical that tests that are capable of detecting false impairments continue to be developed. To that end, a new test of malingered memory performance was researched. The Progressive Visual Memory Test (PVMT) uses a forced-choice (2-choice) paradigm but adds to the literature by including 3-choice and 4-choice trials. A simulation design was conducted in which college students ($N = 62$) acted as uncoached simulators, coached simulators, and controls. A two-way ANOVA explored the impact of group and trial on mean PVMT scores for each trial. The group by trial interaction was not significant. There were statistically significant main effects due to group and trial. Post-hoc analyses revealed significant differences between the control group and the simulator groups on each trial of the PVMT. Controls participants produced almost perfect scores on all three trials while the simulator groups' scores were much lower. Therefore, the PVMT showed excellent sensitivity and specificity. Response latency on the PVMT also differentiated controls from simulators. In addition, when the PVMT was included in a battery of standard neuropsychological tests, participants could not accurately identify the nature of each test nor could they determine which were actual tests of effort, showing that it was difficult to identify the PVMT as an effort test. The PVMT shows excellent promise as an effort test using a new paradigm to identify dissimulation. Limitations and future directions are discussed.

Table of Contents

List of Tables

List of Figures

The detection of exaggerated or malingered symptoms is a complex issue that neuropsychologists regularly encounter during evaluations across a variety of contexts (Bianchini, Mathias, & Greve, 2001; Drob, Meehan, & Waxman, 2009; Greiffenstein, Baker, & Gola, 1994; Mittenberg, Patton, Canyock, & Condit, 2002; Pella et al., 2012; Rees, Tombaugh, Gansler, & Moczynski, 1998; Teichner, & Wagner, 2004). Patients may exaggerate or feign symptoms for an assortment of reasons. For example, those patients involved in litigation or applications for assistance (such disability benefits, etc.) may be motivated to inflate or purposefully exaggerate symptoms, given that this may result in financial compensation (Bianchini, Curtis, & Greve, 2006; Binder, 1993; Ferguson, 2003; Pella et al., 2012; Mittenberg et al., 2002). Patients may also intentionally or unintentionally report symptoms above that which might be expected from an injury for the purpose of meeting psychological needs, as in factitious disorder or conversion disorder (Bush et al., 2005; Drob, et al., 2009; Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009; Mittenberg et al., 2002; Williams & Jones, 2012). It is therefore critical, that neuropsychologists utilize tests and measures that are capable of detecting invalid or false claims of impairment if an accurate diagnosis and appropriate treatment recommendations are to be made. The process by which neuropsychologists identify improbable symptom report or non-credible test performance is called validity testing (Silver, 2015).

Validity testing is a necessary component of all types of neuropsychological evaluations. Once such type of evaluation is the assessment of mild traumatic brain injury, or mTBI. Mild traumatic brain injury accounts for 80% of all TBI cases nation-wide (CDC, 2003), and as such, is one of the most frequently occurring referrals that neuropsychologists encounter (Bigler, 2014; Sohlberg & Mateer, 2001). Unlike moderate or severe brain injuries, where neurological impairments may be more obvious (e.g., difficulty with speech, language, or motor function), the

problems associated with mild TBI are frequently subtle and often involve higher order, executive functions such as paying attention, concentrating, planning, and organizing (Prigatano, 1999). Moreover, patients who report for neuropsychological assessment of mTBI frequently cite symptoms such as mental fatigue, irritability, anxiety, depression, and sleep disturbance-symptoms that coincide with many other conditions. This symptom overlap makes it difficult to differentiate injury-related complaints from complaints that stem from other factors such as life stress or premorbid conditions (Pella, Hill, Ashvind, Hayes, & Gouvier, 2012: Prigatano, 1999). The often-subtle presentation of mTBI is further complicated by neuroimaging studies that have been shown to provide limited diagnostic assistance (Inman & Berry, 2002). Consequently, an accurate evaluation that includes a determination about the impact of symptoms on activities of daily living, can be complex and challenging (Bush, Ruff, Troster, Barth, Koffler, Pliskin, Reynolds, & Silver, 2005; Drob, Meehan, Waxman, 2009; Slick, Sherman, & Berry, 1999). This process can become even more convoluted when complaints occur in the context of external incentives such as financial compensation or applications for disability assistance (Promberger & Marteau, 2013).

Terminology

Although there have been efforts to clearly define validity testing terminology (Slick et al., 1999), there is no current consensus in the literature about *what* to call tests that measure validity of symptoms and/or task abilities, and *how* to describe an improbable performance when it occurs (Bigler, 2012; Bigler, 2014; Bush et al., 2005; Drob et al., 2001; Heilbronner et al, 2009; Larrabee, 2012; Williams & Jones, 2012). In terms of *what* to call validity tests, by convention, the term symptom validity test (SVT) has been used to refer to the large class of tests that measure validity across an assortment of functional domains such as self-reported emotional

and psychological functioning, as well as cognitive functioning (Larrabee, 2012; McCaffrey, Lynch, & Howe, 2012; Pankratz, 1983; Reynolds & Horton, 2012). Yet, use of this term is somewhat imprecise when referring to objective performance on neuropsychological tests; tests that focus on various types of performance *abilities* (Heilbronner et al., 2009; Larrabee, 2012). This indistinct use of language can make it difficult to know exactly what is meant by the term symptom validity test, or SVT. Therefore for the rest of this paper, following the argument put forth by Larrabee (2012), symptom validity testing (SVT) will refer to, "…the accuracy of symptomatic complaint on measures such as the MMPI-2," while performance validity testing (PVT) will refer to, "…the validity of actual ability task performance, assessed either by stand-alone tests such as Dot Counting or by atypical performance on neuropsychological tests such as Finger Tapping…" (p. 626).

In terms of *how* to describe non-credible self-report of symptoms or performance on neuropsychological tests, terms that are commonly used tend to assume the intent behind that report or performance. An extreme example of this is illustrated by the term 'compensation neurosis' which was used in the late 19th and early 20th centuries to describe workers' compensation cases thought to be fabricated (Resnick, 1997). Clearly this term was meant to convey that workers' complaints were purely driven by a desire for monetary compensation. It is likely that some were, but also likely that many complaints were not driven by financial gain. Terms that have been used interchangeably include, 'faking bad,' 'disingenuous performance,' 'exaggerated performance,' or 'feigned symptom reporting' and "malingered performance," along with the more recent trend of referring to 'poor effort,' and/or 'sub- or non-optimal effort' (Bigler, 2014; Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009). A problem with this approach is that these terms have subtle yet distinct meanings and using them interchangeably

loosely implies that they are synonymous when in fact, they are not (Bigler, 2014; Heilbronner et al., 2009; Larrabee, 2012; Slick et al., 1999). Perhaps only the term malingering (from those listed above), may be used with confidence when narrowly and operationally defined as a performance that falls *below* chance responding. A brief description of chance responding itself, will illustrate this point. Consider a test where each item on that test asks the subject to select an answer from two available options. This is called a two-alternative forced-choice (2AFC) paradigm. In this scenario, the subject has two options; one is the right answer and the other is the wrong answer. Thus, the subject has a 50% chance of making a correct choice on any given item by guesswork alone. Stated another way, if a subject guessed on every item of a test that offered two-choice options, over a number of trials, that subject's performance would fall at the 50th percentile, meaning half of the subject's guesses were correct, and the other half of the subject's guesses were incorrect. This is what is meant by chance responding, or that which would occur by guessing alone. Based on this rationale, performances that fall *significantly below* the 50th percentile indicate that something *other than guessing or chance responding has occurred*. Performances that are *significantly below* chance require active decision-making and intentional avoidance of the correct response. In other words, deliberate obfuscation has occurred. Therefore, based on probabilistic analysis, use of the term malingering *is* appropriate in specific cases; however, use of this term is also associated with substantial consequences for false-positive errors (Drob et al., 2009; Pella et al., 2012; Reynolds & Horton, 2012; Slick et al., 1999).

Yet another conundrum includes use of the term 'effort' to describe a necessary, yet indeterminate element or component put forth by individuals during testing. Generally speaking, putting forth effort during testing is thought to provide test data that better represents an

individual's actual abilities. However, discussions in the field of SVT and PVT research are ongoing with respect to an exact definition of the term (Bigler, 2012; Bigler, 2014; Bush et al., 2005; Heilbronner et al., 2009; Larrabee, 2012). For example, Bigler (2012) challenged the notion that effort is a force that is directly and deliberately controlled by each individual over the course of testing. Bigler postulates that the neurobiological networks that underlie "drive" or "effort" can be exactly those which are damaged during injury (2012). It is also now understood that effort is not a static force, that an individual's level of effort, however defined, can vary both over the course of testing and over the course of individual tests (Bigler, 2014; Heilbronner et al., 2009). As such, how should researchers and clinicians characterize the overall performance of someone who "fails" one PVT, but "passes" another? While there is reasonable agreement among clinicians and researchers on the need to administer multiple PVTs (to answer the question above), it is recognized that there is no guaranteed, lock-step way to answer the question of malingered performance. Lastly, how should neuropsychologists interpret two separate PVTs where in one, the subject has failed by a very small margin (i.e., one or two points) and on the other, the subject has failed by a large margin but is still above chance level responding?

At the core of this confusion seems to be the assignment of volition. In other words, are researchers and clinicians able to say with confidence that the individual being evaluated is *deliberately* magnifying symptoms or *purposefully* and *intentionally* avoiding correct responses? Increasingly, research demonstrates that determining intentionality may be possible under specific circumstances such as when multiple sources of data converge in a particular direction (Reynolds, 1998; Reynolds & Horton, 2012; Slick et al., 1999). These sources of data should not be limited to test data, but should also include the clinical interview, behavioral observations,

collateral reports, and educational, medical, and psychological records (Hartlage, 2012; Nies & Sweet, 1994; Reynolds & Horton, 2012; Rogers, 1990).

Commonly Exaggerated or Malingered Sequelae

The list of possible symptoms that may be exaggerated or malingered is lengthy and includes the following list (not exhaustive): 1) cognitive impairments such as diminished attention and concentration, memory loss, and speech and language problems; 2) sensory impairments such as visual field impairments and tinnitus; 3) motor impairments including weakness and slowing; and 4) emotional disturbance such as depression and anxiety (Franzen & Iverson, 1998; Frederick, 2002; Heilbronner et al., 2009; Lynch, 2004; Pella et al., 2012). As mentioned above, exaggerated or malingered symptoms may occur during individual performance validity tests and/or across the entire evaluation in varying degrees (Hartlage, 2012; Hepp, Gamma, Milos, Eich, et al., 2006). To investigate this idea further, Iverson (1995) examined the actual test-taking strategies that were used by a mixed sample of individuals (university undergraduates, volunteers, psychiatric inpatients, and federal prisoners) to malinger during an experimental study. Iverson found that the most frequently reported strategy (16%) was to fake total amnesia (1995), an easily detectable strategy. Some additional strategies included poor cooperation, demonstration of frustration, response latency, and general confusion (Iverson, 1995). Only a small percentage of participants (4%) endorsed utilizing preparation strategies such as learning about the effects of head injury and how best to portray those effects behaviorally (e.g., irritability or distraction). It is critical to note that Iverson completed this work in 1995; a time when use of technology and the internet had not yet become widespread. Today, test security is at a perilous point. For example, a study by Kimpton (2014) that utilized simulated malingerers demonstrated that 71% of participants used the internet to access

information about the neuropsychological test performance of individuals with mild brain injuries in order to better simulate the condition. Further research has demonstrated that locating performance information on the internet (in addition to sensitive information about the objective of a specific test and test-taking strategies) is not difficult (Reynolds, Clark, & Hall, 2017). Given today's accessibility of information via the internet, Iverson's percentages would no doubt climb sharply if that same study were to be administered today.

Base Rates of Malingering

In their influential meta-analysis on malingering across a variety of disorders, Binder and Rohling (1996) point out an interesting and fairly nonintuitive fact about individuals with mild traumatic brain injury. They report that those with mTBI with seek monetary compensation for damages at a higher level than do those with moderate or severe brain injuries (Binder & Rohling, 1996). To be fair, this asymmetry in behavior may not just be related to incentivized mTBI litigants. It may be that those with moderate or severe brain injuries are able settle their legal cases with little to no formal legal dispute (Gouvier, Lees-Haley, & Hayes, 2003; Pella et al., 2004). Despite this fact, it has been demonstrated that individuals with mTBI, who are in litigation, endorse more cognitive complaints than those with mTBI who are not in litigation (Binder & Rohling, 1996; Orey, Cragar, & Berry, 2000). Even more provocative, is the finding that some researchers have described in which a dose-response relationship exists between the magnitude of the monetary incentive and the probability of malingering (Bianchini, Curtis, & Greve, 2006; Larrabee, 2012). These findings are made more striking given the fact that most individuals make a complete recovery from mTBI in three months or less (Ponsford, Cameron, Fitzgerald, Grant, Mikocka-Walus, & Schonberger, 2012).

Observations such as those above compel the question: what are the base rates of malingering in various disorders and conditions that are routinely seen at evaluation? Unfortunately, malingerers are unlikely to reveal to neuropsychologists that they are malingering (Hartlage, 2012), even when they are confronted with evidence that clearly demonstrates below chance responding (Babin & Gross, 2002; Langeluddecke & Lucas, 2003). Mittenberg, Patton, Canyock, and Condit (2002) systematically pursued this central question in their extensive review entitled, "Base Rates of Malingering and Symptom Exaggeration," published in the *Journal of Clinical and Experimental Neuropsychology*. A number of meaningful findings arose from their work. First, base rates of malingering and exaggeration vary with different diagnoses and vary according to context (Mittenberg, Patton, Canyock, & Condit, 2002). For example, irrespective of the referral source (and using 95% confidence intervals) the following mean base rates of probable malingering were observed for groups of individuals with the following diagnoses: fibromyalgia or chronic fatigue (41.24), pain or somatoform disorders (33.50), depressive disorders (16.08), anxiety disorders (13.57), and moderate or severe brain injury (8.82). Interestingly, the mean base rate of probable malingering in mildly brain injured patients was found to be 41.24, a much larger base rate than that found in more severe cases of brain injury (Mittenberg et al., 2002).

With respect to the differing contexts in which probable malingering occurs, Mittenberg et al. (2002) found that the mean base rate was 30.43 for personal injury cases, was 32.73 for disability or worker's compensation cases, was 22.78 for criminal cases, and was 8.11 for medical or psychiatric cases (Mittenberg et al., 2002). It is possible that these context-related mean base rates may be due, in part, to the effects of coaching. For example, Essig, Mittenberg, Petersen, Strauman, and Cooper (2001) note that it is common practice for attorneys to prepare

their clients for neuropsychological examinations on such features as expected symptoms given degree of injury, descriptions of actual tests, and methods used in the detection of malingering. It is important to note that the mean base rates provided by the Mittenberg et al. analysis (2002), were all sourced from research that adhered to the criteria for Malingered Neurocognitive Dysfunction (MND) put forth by Slick, Sherman, and Iverson (1999).

Slick Criteria for Malingering of Neurocognitive Dysfunction (MND)

Malingering has been described in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) (American Psychiatric Association [APA], 2013); however, descriptions are brief and primarily convey the option of assigning a V-code to an evaluation that includes a malingered test performance. A V-code of malingering provides limited utility to neuropsychologists who must carefully dissect each test performance and then offer an overall opinion that captures all of the data (test and non-test) for each evaluation. Malingering is one option among many that may be used to describe an improbable response style; therefore neuropsychologists need a variety of terms to reflect a spectrum of possible performances. Despite this limitation, the information included in the DSM-V does offer important guidance about differentiating malingering from factitious disorder, conversion disorder, and other somatic symptom-related disorders by examining other factors such as external versus internal incentives (APA, 2013). Fortunately, an in-depth, well-defined rubric for determining malingering was published nearly two decades ago.

Arguably one of the most important papers in the discussion of malingering was put forth by Slick, Sherman, and Iverson in 1999. Although numerous researchers have contributed in varying and significant ways to the identification of criteria for assessing malingering (Rogers, 1990; Greiffenstein, Baker, & Gola, 1994; Pankratz and Binder, 1997; Faust and Ackley, 1998;

Nies and Sweet, 1994). Slick et al. (1999) expanded the literature by proposing a systematic approach to malingering identification. Slick and colleagues advised the use of four primary criteria (each with corresponding sub-criteria) and three degrees of diagnostic certainty when determining malingered neurocognitive dysfunction (MND); a disorder the authors argue warrants its own diagnostic label. Slick et al. define MND as follows:

> "Malingering of Neurocognitive Dysfunction (MND) is the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain or avoiding or escaping formal duty or responsibility. Substantial material gain includes money, goods, or services of nontrivial value (e.g., financial compensation for personal injury). Formal duties are actions that people are legally obligated to perform (e.g., prison, military, or public service, or child support payments or other financial obligations). Formal responsibilities are those that involve accountability or liability in legal proceedings (e.g., competency to stand trial)" (p. 552).

Degrees of diagnostic certainty receive labels of Definite, Probable, or Possible based on the extent to which the four primary criteria (and their sub-criteria) are met. Primary criteria include: A) Presence of a substantial external incentive; B) Evidence from neuropsychological testing; C) Evidence from Self-Report; and D) Behaviors meeting necessary criteria from groups B and C are not fully accounted for by Psychiatric, Neurological, or Developmental Factors (Slick et al., 1999).

Slick, Sherman, and Iverson (1999) put forth clear, consistent, and well-defined standards for use in identifying malingering. Slick and colleagues encourage neuropsychologists to use these standards consistently in clinical practice and in research so that the field may advance uniformly. They also note that these standards should not be applied in an inflexible or rigid

fashion. Slick et al. recognize that each case is comprised of a unique constellation of features that ultimately requires the use of clinical judgement in order to assemble the pieces in the most reasonable and probable manner (1999).

Formal Evaluation of Malingering

In 2009, the American Academy of Clinical Neuropsychology (AACN) published a consensus conference statement on neuropsychological assessment of effort, response bias, and malingering (Heilbronner, Sweet, Morgan, Larrabee, & Millis). This comprehensive document offers a rich body of information ranging from definitions of terms such as malingering "detection," versus "diagnosis," effort and intentionality, to types of assessment methods, types of inconsistencies seen during evaluation, and recommendations for assessments according to symptom type (2009).

According to Heilbronner et al., neuropsychologists have at their disposal a wide variety of tools from which they may derive explicit statements about malingering (2009). Important sources of data include the clinical interview, behavioral observations, collateral reports, educational data, and medical or psychological records (Hartlage, 2012; Nies & Sweet, 1994; Reynolds & Horton, 2012; Rogers, 1990; Slick et al., 1999). In addition to these data are data gleaned from measures of self-report such as disorder-specific inventories (e.g., PTSD, eating disorders, phobias), or psychopathology and personality measures such as the Minnesota Multiphasic Personality Inventory – Second Edition (MMPI-II) (Butcher, Dalstrom, Graham, et al., 1989), the Personality Assessment Inventory (PAI) (Morey, 1991), and the Millon Clinical Multiaxial Inventory – Third Edition (MCMI-III) (Millon, 1977). Heilbronner et al. strongly underscore the need for such inventories by stating the following: "No examiner in any discipline is required to accept self-reported facts and history of examinees. The validity of self-reported

disability and symptoms needs to be evaluated, especially when such complaints occur in a forensic context" (p. 1102). Comparison of the aforementioned types of data (clinical interview, behavioral observations, self-report, etc.) with data collected via performance validity testing (PVT) is essential if one is to confront the possibility of malingering (Heilbronner et al., 2009). General characteristics of invalid presentations include: presentations that are not fully explained by brain dysfunction, presentations that are not reasonably attributable to moderating and/or confounding variables (e.g., education and fatigue), and presentations that are worse than the performance of known groups with bona fide neurological disorders (Heilbronner et al., 2009).

Types of Performance Validity Tests

The late 1970s and early 1980s saw a rapid expansion of research focused on validity testing (Bigler, 2012; Bush et al., 2005; Heaton et al., 1978; Heilbronner et al., 2009; Lynch, 2004). Prior to that time, clinical judgement was the primary tool by which assessments of symptom and performance validity were made (Heaton et al., 1978; Reynolds, 1998; Reynolds & Horton, 2012). It was mistakenly thought that clinical experience alone had the power to accurately and consistently detect feigned symptoms (Slick, Sherman, & Iverson, 1999). Since then, consensus in the field is that this is definitely not the case (Reynolds, 1998; Reynolds & Horton, 2012; Slick, Sherman, & Iverson, 1999). Research now demonstrates that clinical judgement is notoriously inaccurate (Bigler, 2012) and when used alone achieves only chance level accuracy (Heaton, Smith, Lehman, & Vogt, 1978; Lynch, 2004). This is not to say that there is no place for expert clinical judgement; however, expert clinical judgement should arise from multiple data sources, not the least of which is the test data.

Traditional neuropsychological PVTs may be categorized as either stand-alone or embedded measures (Bigler, 2014; Erdodi & Lichtenstein, 2017; Martin, Schroeder, & Odland,

2015). Stand-alone measures are measures that were developed specifically to evaluate the validity of task performance (Boone, 2013; Green, 2013; Sollman & Berry, 2011) and they are now deemed to be medically necessary (Bush, Ruff, Troster, Barth, Koffler, Pliskin, Reynolds, & Silver, 2005). Stand-alone PVTs may be selected for administration based on the reported type of cognitive complaint (Lynch, 2004; Slick et al., 1999); however, most often PVT selection is based on the judgement of the clinician (Bigler, 2012). For instance, PVTs that have been created to evaluate the validity of different types memory performance include the Portland Digit Recognition Test (PDRT) (Binder, 1993), the Test of Memory Malingering (TOMM) (Tombaugh, 1996), and the Word Memory Test (WMT) (Green, 2003). Other domains that may be evaluated include motor skill performance (Larrabee, 2003), and sensory/perceptual abilities (Pankratz, 1979).

PVTs often utilize known groups (KG) design. In KG design, the clinician or researcher compares a subject's score on a PVT to that of different groups of individuals with bona fide neurological impairments (Greve & Bianchini, 2004; Larrabee, 2012). PVTs such as this employ a clear-cut, straightforward task that is simple and does not challenge even individuals with demonstrated memory impairments such those with Alzheimer's disease. Research has shown that individuals with demonstrated memory impairments are able to achieve perfect or near-perfect performances on PVTs (Binder, 1993; Green, 2003; Tombaugh, 1996). Thus, when an individual with less significant impairments scores below those with more significant impairments, alternative explanations for that performance must be considered (Vickery, Berry, Inman, Harris, & Orey, 2001). It is important to note in this discussion that PVT "failure" is not synonymous with malingering (Bigler, 2014; Bush et al., 2005; Heilbronner et al., 2009). It is only when an individual's performance falls *significantly below* the performance of those with

bona fide neurological conditions (and other conditions are met [Slick et al., 1999]) is malingering thought to be definite. This is the strength of the KG design; given the performance of individuals with severe neurologic dysfunction, a poorer performance by a more neurologically intact individual suggests deliberate or intentional avoidance of correct responses (Bianchini, Mathias, & Greve, 2001; Reynolds & Horton, 2012; Slick, 1999).

Embedded measures are indicators that are found within other tests of ability (Schwartz, Erdodi, Rodriguez, Ghosh, Curtain, Flashman, & Roth, 2016) and are assumed to provide validity information specifically about the neuropsychological domain to which the parent measure belongs (Greve, Bianchini, Mathias, Houston, & Crouch, 2003; Pella et al. 2012). Commonly used embedded measures include tests such as Reliable Digit Span (Greiffenstein, Baker, & Gola, 1994) which is derived from the Digit Span subtest of the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV) (Wechsler, 2008), and the Recognition task found within the California Verbal Learning Test – Second Edition (CVLT-II) (Donders & Strong, 2011; Schwartz, Erdodi, Rodriguez, Ghosh, Curtain, Flashman, & Roth, 2016). Advantages to the use of embedded measures include time saved in test administration (Barrash, Suhr, & Manzel, 2004), embedded measures are less susceptible to coaching, are less identifiable as PVTs, and can provide information about performance validity across the entirety of the evaluation (Pella et al., 2012).

It is strongly recommended that multiple PVTs, both stand-alone and embedded measures, be utilized in any one evaluation, as, "…no single finding is sufficient to identify malingering" (Pella et al., 2012, p.128). Simply put, the use of stand-alone and embedded measures together, yield more data points from which a malingering determination can be made. Failure on multiple PVTs greatly increases the probability of malingering, specifically in the

context of external incentives (Larrabee, 2012). For example, Larrabee (2003) demonstrated that failure on two embedded PVTs and/or SVTs resulted in a sensitivity of .875 and specificity of .889.

Diagnostic Accuracy of PVTs: Sensitivity, Specificity, and Predictive Power

Bianchini and colleagues (2001) assert that the ultimate measure of success of tests of malingering is the ability to accurately classify individual patients. The indices by which neuropsychologists make these classifications are sensitivity, specificity, and predictive power (Bianchini, Mathias, & Greve, 2001; Larrabee, 2012; Lynch, 2004). These indices can be shaped by a number of factors including features of the population or sample of interest, features of the selected PVT, and other factors such as coaching and/or the patient's knowledge about the effects of TBI (Bianchini, Mathias, & Greve, 2001; Mittenberg et al., 2009; Hennekens & Buring. 1987). Simply stated, sensitivity is the true positive rate for a test, or the number of people with a condition who had a positive result divided by all the people with the condition, whereas specificity is the true negative error rate, or the number of people without a condition who had a negative result divided by all the people without the condition (Bianchini, Mathias, & Greve, 2001; Heilbronner et al., 2009; Lynch, 2004). Instruments are frequently selected based on sensitivity and specificity; an instrument should have few false negatives (it should consistently identify malingerers) as well as few false positives (it should not identify normals as malingerers) (Lynch, 2004). It is typical for researchers in the field to set specificity for individual tests at as close to 90% as possible. Keeping specificity as high as possible allows for fewer false positive errors (Boone, 2013; Larrabee, 2012). While important components in test selection, sensitivity and specificity are independent of base rates; data which are extremely helpful in generalizing group research to individual cases (Gouvier, 1999; Larrabee, 2012; Pella

et al., 2012). It may be that PVTs with high sensitivity and specificity do not outperform base rate information alone in populations with small base rates. Because clinicians and researchers want to use PVTs that are *better* at identifying malingering than base rates alone, they may determine the predictive power (positive and negative) of a PVT. Predictive power is independent of sensitivity and specificity. Positive predictive power (PPP) increases when the behavior of interest (i.e., malingering) increases, while negative predictive power (NPP) decreases when the behavior of interest decreases. In this way, the predictive value of a PVT is influenced by factors that affect base rates (e.g., litigants v. non-litigants, forensic context v. medical context) in addition to cut-off scores (Baldessarini, Finklestein, & Arana, 1983; Bianchini, Mathias, & Greve, 2001; Mittenber et al., 2002; Pella et al., 2012). Therefore, PPP indicates the probability that a person *does* have a condition given a *positive* test result, while NPP indicates the probability that a person *does not* have a condition given a *negative* test result (Bianchini, Mathias, & Greve, 2001; Pella et al., 2012). PPP and NPP act like barometers of confidence; in other words, they represent the degree to which one can be confident that the determination made (based on the score that was obtained) is accurate (Bianchini, Mathias, & Greve, 2001). The closer PPP and NPP are to 100, the more confidence we can have that we are making accurate determinations regarding malingering.

Brief History of the Forced-Choice Paradigm

The first forced-choice paradigms were described in the literature as tools in the detection of malingering in the late 1970s and mid-1980s (Binder & Pankratz, 1987; Pankratz, 1979). As mentioned, the foundation of this approach is a two-alternative forced-choice (2AFC) test that results in a known level of chance responding (50%), whereby a score that falls *below* chance responding indicates active avoidance of correct responses (Bianchini, Mathias, & Greve, 2001;

Reynolds & Horton, 2012). Indeed, Hiscock and Hiscock (1989) highlight the value of this statistical fact with the following assertion from their influential paper on 2AFC tests, "A person must be capable of scoring significantly above chance in order to score significantly below chance," (p. 968).

Hiscock and Hiscock (1989) built upon the method used by Binder and Pankratz (1987) whereby a woman was asked to specify whether a yellow pencil or a black pen had been presented on 100 trials. Her performance fell significantly below chance at 37% at the $p < .01$ level, making it clear that there had been intent to deceive. Hiscock and Hiscock recognized that the Binder and Pankratz procedure might not be effective with higher functioning individuals as it had little to no face validity as an actual test of cognitive functioning (1989). Hiscock and Hiscock correctly hypothesized that the PVT must appear to the test-taker to be a legitimate test of cognitive ability (i.e., a test of memory), thereby avoiding suspicion or detection of the true purpose of the PVT by the test-taker. Consequently, Hiscock and Hiscock changed the stimuli to a 5-digit string of numbers and varied the interstimulus interval (ISI) which together lent an increased perception of task difficulty without increasing the actual difficulty of the task. Experimental results were similar to those found by Binder and Pankratz; the patient's score fell significantly below chance at 29%, while control subjects scored much higher (Hiscock and Hiscock, 1989).

Since that time, PVT test development has greatly increased (Heilbronner et al., 2009), and the most well-validated, peer-reviewed instruments still employ a forced-choice paradigm (Binder, 1993; Green, Lees-Haley, & Allen, 2002; Tombaugh, 1995). Variations on this theme include presenting visual stimuli such as the simple line drawings found in the Test of Memory Malingering (TOMM) (Tombaugh, 1995), or the numbers, letters, and shapes found in the Rey

15-Item Memory Test Plus Recognition Trial (Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002; Rey, 1941). While PVTs provide indispensable information to the researcher or clinician, they are not infallible tools. As previously stated, stand-alone PVTs can be vulnerable to the effects of coaching whereby litigants may be instructed on ways to mimic poor performance without being detected (Hausknecht, Halpert, DiPaolo, & Gerrard, 2007; Mittenberg et al., 2002; Pella et al., 2012; Youngjohn, 1995). It is not uncommon for attorneys to advise their clients about what information to disclose regarding their injury and what information not to disclose (Essig, Mittenberg, Peterson, Strauman, & Cooper, 2001). Research has clearly demonstrated that PVT test-taking strategies and actual stimuli can be easily accessed via the internet and with minimal skill (Reynolds, Clark, & Hall, 2017; Kimpton, 2014). Therefore, based on these observations, it is possible that tests may wane in their ability to accurately identify malingering over time. Test security has reached a critical level. Without new and novel measures of test validity, Type II error rate will increase to unacceptable levels.

The Progressive Visual Memory Test (PVMT)

The Progressive Visual Memory Test (PVMT) is a new, computerized, 45-item, stand-alone PVT designed to detect malingered memory performance. The PVMT uses a two-alternative forced-choice paradigm; however, it expands the literature in a novel way: the PVMT also includes 3-choice, and 4-choice trials. Meaning, on the first trial the participant must choose between two stimuli, on the second trial the participant must choose between three stimuli, and on the third trial the participant must choose between four stimuli. The purpose of this addition is twofold. First, increasing the number of stimuli per item may serve to enhance the face validity of the PVMT as a test of memory, without negatively impacting patients' performance (Hiscock & Hiscock, 1989). Research has shown that tests that appear overwhelmingly easy (especially

when compared to other more challenging neuropsychological tests) may raise participants' suspicions about the nature of the test and ultimately influence performance (Hiscock & Hiscock, 1989). Therefore, it is critical that the test feel realistic to the test-taker, ergo, the increase in stimuli. Second, the addition of 3-choice and 4-choice trials could lead to two interesting outcomes for neurologically normal individuals. One outcome may be that all trials for each neurologically normal individual result in perfect or near-perfect scores, indicating that the PVMT successfully functioned as expected. In other words, the PVMT did not actually present a memory challenge, largely due to the fact that individuals have a high capacity for storing and retrieving complex visual information (Rees, Tombaugh, Gansler, & Moczynski, 1998). This result may be also be facilitated by the fact that the same 45 target stimuli will be repeated on each of the three trials, thus greatly increasing participants' familiarity with the target stimuli. Alternatively, it may be that trial 2 and /or trial 3 present a legitimate memory challenge for neurologically normal subjects. If this is the case, the PVMT may transition to an actual memory test at trial 2 or at trial 3. This result would potentially double the utility of the PVMT. This is an approach that was taken on the Word Memory Test (WMT; Green, 2003), where both validity and memory indices are generated. The WMT achieves this by changing methodology between the validity and memory indices, i.e., starting with a forced-choice paradigm and converting to commonly used memory paradigms such as multiple choice, paired associates recall, and delayed free recall trials. In contrast, the PVMT may offer both validity and memory indices without the necessity of changing methodology during the test, i.e., using a forced-choice paradigm across all trials. Either of the results described (a new stand-alone PVT or a new stand-alone PVT that transitions to a test of memory) add novel components and a novel approach to the existing body of research on PVTs.

The performance characteristics of the PVMT will be explored with a sample of college students participating in research as a part of an introductory psychology course at a Northwestern university. A simulation design will be employed where subjects will be semi-randomly assigned to one of three groups: uncoached simulators, coached simulators, and control subjects. The primary aim of the study will be to administer the PVMT with neurologically normal individuals (control group), putting forth their best effort, and to observe how the test operates. The secondary aim of the study will be to observe the performance of simulated malingerers, where one group is coached on how to portray negative cognitive symptoms in a realistic and subtle manner (coached simulators) and the other group is not (uncoached simulators).

Hypotheses by group:

1) It is hypothesized that mean PVMT scores for coached simulators will be significantly lower than scores for the control group on each trial of the PVMT, where the null hypothesis is $H_0 = \mu_{coached} = \mu_{control}$.
2) It is hypothesized that mean PVMT scores for uncoached simulators will be significantly lower than scores for the control group on each trial of the PVMT, where the null hypothesis is $H_0 = \mu_{uncoached} = \mu_{control}$.
3) It is hypothesized that mean PVMT scores for uncoached simulators will be significantly lower than scores for coached simulators on each trial of the PVMT, where the null hypothesis is $H_0 = \mu_{coached} = \mu_{uncoached}$

Hypotheses across trials:

4) It is hypothesized that there will be an interaction between group and trial such that

A. the uncoached simulators will decrease at a significantly greater rate than the coached simulators across the three trials, where the null hypothesis is

$H_0 = \mu_{uncoached} = \mu_{coached}$, and

B. the coached simulators will decrease at a significantly greater rate than the control group across the three trials, where the null hypotheses is

$H_0 = \mu_{coached} = \mu_{control}$.

Method

Participants

Undergraduate students from the University of Montana were recruited for participation from an introductory course in psychology. Participants were recruited during a designated screening day where they completed an informed consent form and a demographic questionnaire (see Appendices A and B) to determine if they were eligible to participate. Students were not asked to participate if they were younger than 18 years of age. In order to obtain a sample of participants without neurological or psychiatric impairments, individuals were not invited to participate if they reported birth difficulties, learning difficulties, neurological conditions, current treatment for psychological conditions, or endorsed more than two out of five items that assessed alcohol or drug abuse. Students who met inclusion criteria were contacted by the principal investigator and invited to participate. Sixty-two participants completed the entire study and all data were included in the analysis.

All participants received four credits that were applied toward an experimental credit requirement in their psychology course. Participants were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2002). In order to protect confidentiality, participants' names and any other identifying information was

detached from any data they provided. Participants were assigned an identification number that was used to keep participants' data organized.

Measures

A note regarding measures. The test battery for this study was composed of the primary measure of interest (PVMT) as well as five additional, commonly administered neuropsychological tests. Only data from the PVMT were used for primary analyses. Other tests were added to the battery to provide data for exploratory analyses and to simulate more realistic testing conditions.

Informed consent form (ICF). The ICF (Appendix A) for the PVMT study was completed on screening day and provided potential participants with information about the study including the name and contact information of the principal investigator, study purpose, expected costs and benefits to participation, indicated that participation would yield 4 research credits, and would take 60 minutes to complete. The ICF also indicated that participants could withdraw at any point without penalty. Participants who did not sign the form were not invited to participate.

Demographic questionnaire (DQ). The DQ (Appendix B) was also completed on screening day and was used to gather participant characteristics including age, gender, race, and education. The DQ was administered to screen participants and exclude those who endorsed neurological impairments, those receiving treatment for psychological symptoms, those with a history of TBI, and those who endorsed problems with alcohol or drug use (positive endorsement of greater than two out of five such items and/or endorsed current treatment for a psychological condition).

Group assignment with corresponding instructions. At the study appointment, each participant received an envelope containing pseudo-random assignment to one of three groups:

uncoached simulators, coached simulators, or control participants. Instructions for participation varied with group assignment (see Appendices C, D, and E) and were coded such that both the participant and the researcher remained blind to group assignment.

Progressive Visual Memory Test (PVMT). The PVMT is a computerized test that consists of three, forced-choice trials that contain 45 items per trial, for a total of 135 items. Prior to the start of the first trial, there are instructions followed by two practice items. The instructions are displayed on the computer monitor and read aloud by the researcher. The instructions are as follows:

> *"This is a test of your ability to learn and remember pictures. First you will be shown a series of pictures, on at a time. Then, you will have a chance to see how many of them you can remember. Do you best to remember each picture."*

Next, a screen displays the following:

> *"Let's try a sample that contains two pictures. Look carefully at each picture and try to remember it. You don't have to learn the name of each picture. Just look at each one and try to remember it."*

The following screen commences the learning trial and displays a target stimulus (i.e., a red apple) for three seconds, followed by a black screen for one second which serves as the interstimulus Interval (ISI). This pattern repeats again (3-sec stimulus/1-sec ISI) with another unique target stimulus (i.e., a white bird). Immediately following the presentation of the two target stimuli (red apple and white bird), the recognition phase begins whereby two stimuli are presented together (one stimulus that was seen before and one novel *but related* stimulus, i.e., red apple and bowl of red apples). The subject is then asked to indicate which of the two stimuli (A or B) was seen earlier. The subject provides a verbal response, the researcher notes the

response on the answer sheet, and advances the presentation to the second recognition stimulus containing two images (one stimulus that was seen before, and one new *but related* stimulus, i.e., white bird and brown bird). Again, the subject is asked to identify which of the two stimuli (A or B) was seen earlier and the researcher records the response on the response sheet. Each target stimulus always appears as an option on its corresponding recognition trial. Subjects are prompted to respond after 20 seconds (Tombaugh, 1996). If an error is made on a practice item (only), the researcher returns to the original target image, points out the correct image, and confirms that the participant understands how to participate. Once the practice items are complete, the subject reads a screen that asks if he or she has any questions before moving on to Trial 1. The Trial 1 instructions are as follows:

> *"Now you will be shown a series of 45 pictures, one at a time. Look at each one and try to remember it so you can recall it later. Do you best, even if it is hard for you."*

Trial 1 begins, and each stimulus is shown for three seconds with a 1-sec ISI. Following presentation of all 45 stimuli, the Trial 1 recognition phase begins. The researcher records the subject's responses *without* providing verbal feedback to the subject on whether he or she was correct as some research suggests this produces more sophisticated malingering (Youngjohn, Lees-Haley, & Binder, 1999). Responses are noted on a record response sheet and then the researcher moves on to Trial 2. The Trial 2 instructions are as follows:

> *"You will now be shown the same 45 pictures again. Look at each one and try to remember it. On <u>this</u> trial, you will then identify each picture from <u>two</u> other pictures. This trial may be more difficult than the previous trial. Again, try to remember each picture. Just do your best, even if it is hard for you."*

Trial 2 begins, and each stimulus is shown for three seconds with a 1-sec ISI. Following presentation of all 45 stimuli, the Trial 2 recognition phase begins, and the researcher records responses, and moves on to Trial 3; the last trial. The Trial 3 instructions are as follows:

> *"You will be shown the same 45 pictures again. Look at each one and try to remember it. On this trial, you will then identify each picture from three other pictures. This trial may be more difficult than the previous trial. Again, try to remember each picture. Just do your best, even if it is hard for you."*

Trial 3 begins, and each stimulus is shown for three seconds with a 1-sec ISI. Following presentation of all 45 stimuli, the Trial 3 recognition phase begins, and the researcher records responses. The test ends after the Trial 3 recognition phase and three scores are generated for each subject: A Trial 1 total correct score, a Trial 2 total correct score, and a Trial 3 total correct score. This will allow the researcher to analyze differences between groups at the level of each trial. The PVMT takes approximately 20 minutes to complete. For technical specifications (photo sizes, hardware specifications, ppt specifications), see Appendix J.

Word Reading subtest of the Wide Range Achievement Test – Fourth Edition (WRAT4; Wilkinson & Robertson, 2006). The Word Reading subtest (adult level) of the WRAT4 begins with a 55-word reading and pronunciation list where each stimulus is given a 10-second time limit (Wilkinson & Robertson, 2006). The test is discontinued after ten failures and normative information spans ages 5 to 94. The Word Reading subtest of the WRAT4 assumes that familiar words will be pronounced correctly, with more correct words indicating a higher vocabulary level (Wilkinson & Robertson, 2006). Measures of vocabulary have been shown to correlate strongly with intellectual ability (Heaton, Ryan, & Grant, 2009; Yates, 1954). The

Word Reading subtest of the WRAT4 was administered to provide data for exploratory analyses that investigated the performance characteristics of simulators across the test battery.

Trail Making Tests A and B (TMT A and TMT B; Halstead-Reitan Battery [HRB]; Reitan & Wolfson, 1993). Trail Making Tests A and B are tests of scanning, visuomotor tracking, and speed (both Trails A and B) as well as divided attention and executive functioning (Test B). In Trail Making Test A, subjects are asked to consecutively connect numbered circles spread disparately across a page. In Trail Making Test B, subjects are asked to follow the same process except that they must alternate connecting numbers with letters. Performances of patients with mTBI have been shown to be slower than controls and slowing increases with injury severity (Lezak et al., 2012). Dikmen, Heaton, Grant, & Temkin, (1999) demonstrated test-retest reliability as moderate for both Part A (r = .70) and Part B (r = .89). Trail Making Tests A & B have been shown to have low specificity, but high sensitivity, which suggests that they are effective in detecting the presence of deficits, but ineffective in identifying specific deficits (Cicerone & Azulay, 2002). As stated above, TMT A and TMT B were administered to provide data for exploratory analyses that investigated the performance characteristics of simulators across the test battery.

Test of Memory Malingering (TOMM; Tombaugh, 1996). The TOMM is a well-validated, stand-alone PVT that is cited as one of the most frequently used PVTs (Martin, Schroeder, & Odland, 2015; Slick, Tan, Strauss, & Hultsch, 2004). The TOMM contains two learning trials where each trial is followed by recognition memory testing. There is an optional retention trial which was not used in the current study. The learning trials consist of the serial presentation of 50 stimuli (simple line drawings), each for three seconds. Then, the subject is asked to identify from two stimuli, which stimulus they saw before, during the serial

presentation. Cut-off scores are derived from important clinical groups such as Alzheimer's groups and TBI groups. Chance responding is known to be 50%; therefore, below chance responding indicates active avoidance of correct responses. Although the TOMM is a helpful tool in the determination of malingering, increasing numbers of litigants are now familiar with the TOMM or have been coached on how to perform (Lezak et al., 2012). The TOMM has been shown to have high specificity and high sensitivity when used in the detection of malingering and other forms of suboptimal performance (e.g., Tombaugh, 1997; Rees, Tombaugh, Gansler, & Moczynski, 1998; Vallabhajosula & van Gorp, 2001). Moreover, scores do not appear to be impacted by age, education, or neurological disorders (Rees, Tombaugh, & Boulay, 2001; Rees et al., 1998; Tombaugh, 1997). The TOMM was administered to provide data for exploratory analyses that include a comparison of the sensitivity and specificity of PVMT and TOMM.

Symbol Cancellation subtest of the Kaplan Baycrest Neurocognitive Assessment (KBNA™; Leach, Kaplan, Rewilak, Richards, & Proulx, 2000). The KBNA battery is comprised of seven subtests that were adapted from current measures and assessment techniques (Lezak et al., 2012). Symbol Cancellation is a KBNA subtest that consists of a single page with many, small symbols scattered across it in a random fashion. A target shape is printed at the top of the page. The subject is instructed to draw a line through all shapes on the page that match the target shape as quickly as possible. Symbol Cancellation is a quick, simple, and well-validated measure of attention. Symbol Cancellation was administered to provide data for exploratory analyses that investigated the performance characteristics of simulators across the test battery.

Aphasia Screening Test (AST; Halstead & Wepman, 1959). The original AST contained 51 items and screened for language and communication problems. Halstead and Reitan reduced the number of items to 32 and included the AST in their Halstead-Reitan Battery (HRB).

Because of its inclusion in the HRB, the AST is one of the most widely used screening measures for aphasic disorders. During the AST, the subject is asked to verbally identify, spell, and copy common shapes as well as read short passages, repeat phrases, and complete simple computation problems. As above, the AST was administered to provide data for exploratory analyses that investigated the performance characteristics of simulators across the test battery.

Role Play Termination (RPT) Instructions. RPT instructions (Appendix F) were read by each subject once testing was complete. These instructions directed subjects to cease all role play and answer subsequent post-test questionnaires honestly and according to subjects' own views.

Post-Test Questionnaire 1 (PTQ1). The PTQ1 (Appendix G) is a brief questionnaire that first asked participants to summarize the instructions they received at the start of the experiment. Then participants indicated on Likert scales (1-5, with 5 being the highest rating) how hard they tried to simulate the conditions outlined in the instructions, and how successful they thought they were in following the instructions. Lastly, participants were asked to indicate what they thought each test in the battery was designed to measure. Data collected were used in exploratory analyses.

Post-Test Questionnaire 2 (PTQ2). The PTQ2 (Appendix H) explained that some of the tests that participants took were designed to detect when someone feigns brain damage, while others were routine tests used to assess cognitive functioning. The PTQ2 asked participants to make a mark next to any test they thought was designed to detect someone feigning brain damage. Then, participants were asked to indicate on a Likert scale, their degree of certainty that the test they indicated was a test designed to detect someone faking brain damage. Data collected were used in exploratory analyses.

Debriefing Statement (DS). The DS (Appendix I) was given to each participant after completion of the post-test questionnaires. The DS reviewed the purpose of the study, asked subjects to keep all study activities confidential, and thanked subjects for participation.

Procedure

At the study appointment, the researcher and the subject first reviewed informed consent. Then the researcher would hand the participant an envelope and indicate that it contained instructions for participation in the study. The researcher would direct the participant not to open the envelope until the researcher had stepped out of the room, thereby allowing the researcher to remain blind to the conditions of participation for each subject. Once the researcher had stepped out of the room, the subject would open the envelope and read the instructions. The instructions for each participant varied based on pseudo-random group assignment to one of three groups: coached simulators, uncoached simulators, or control group. Subjects were not aware of the group to which they were assigned. Participants in both the uncoached simulator and coached simulator conditions read a vignette depicting a driver who sustained a concussion following a slow-speed motor vehicle collision. Participants in the uncoached simulator condition were instructed to complete all subsequent tests as if they themselves were the injured party from the vignette; an individual who sustained a concussion (Appendix C). Participants in the coached simulator condition were instructed to complete testing in a similar fashion, however, additional details about cognitive functioning after concussion were provided *to this group only* (Appendix D). This information was given to the coached simulator group to assist them in more skillfully simulating the test performance of someone with a concussion. Participants in the control condition did *not* read the vignette depicting the motor vehicle collision; rather, they were instructed to complete all testing to the best of their ability (Appendix E).

To encourage subjects in the uncoached and coached conditions to put forth their best effort in their task of portraying mTBI, subjects in these two groups read an additional prompt on their instruction sheet. This prompt indicated the possibility of earning two additional research credits (for a total of four credits) by performing in such a way that their simulation would be undetectable. It was predicted that the potential for earning a total of four credits would be perceived by participants as a large benefit. This was due to the fact that the population from which the current sample was drawn (introductory psychology students) were required to earn eight research credits across the semester through research participation. Control subjects received an alternative prompt that indicated the possibility of earning two additional research credits by putting forth consistent effort across all tests. Thus, for all groups, earning four credits for participation in a single study was expected to strongly incentivize participants. Ultimately, all subjects received a total of four credits for participation regardless of whether simulation was detectable.

Once the instructions were read, participants were directed to place them back in the envelope, seal the envelope, and place an 'X' across the seal. The participant would then alert the researcher that he or she was ready to begin participation.

Next, participants completed the PVMT within a battery of commonly used neuropsychological tests. A Latin Square design was used to prevent effects due to order. The tests included:

1. Progressive Visual Memory Test (PVMT)
2. Word Reading subtest of the WRAT4 (Wide Range Achievement Test)
3. Trail Making Tests A and B (TMT A and B)
4. Test of Memory Malingering (TOMM)

5. Cancellation subtest of the Kaplan Baycrest Neuropsychological Assessment (KBNA)
6. Aphasia Screening Test (AST)

Following completion of all testing, participants received Role Play Termination Instructions that instructed them to cease all role play. Participants were then given Post-Test Questionnaire 1 (PTQ1) which was completed and collected before Post-Test Questionnaire 2 (PTQ2). Lastly, all participants received a debriefing statement and any questions were addressed by the researcher. Participation lasted one hour.

Data analysis. A 3 groups (coached, uncoached, and controls) x 3 trials (Trial 1, Trail 2, and Trial 3) analysis of variance (ANOVA) was conducted to explore the impact of group and trial on mean PVMT scores. In all analyses, data were inspected for violations of the assumptions required for ANOVA. In the following analyses, the assumption of homogeneity of variances was frequently violated. This is not uncommon in simulation research; simulators often demonstrate greater variability in scores than do the control group (Martin, Hayes, & Gouvier, 1996). As a general rule, if the number of subjects in each group is equal and the variances are no greater than 5 or 6 times each other, the *F* ratio can be interpreted without correction (Howell, 2010). For primary analyses, when the variances were greater than 6 times each other, or when the significance level for Levene's statistic was below .05, a more stringent significance level (.01) was set to reduce Type I error rate (Pallant, 2005). When the assumption of sphericity was not met, the Greenhouse-Geisser statistic was used as a corrected statistic to interpret the F ratio. When the Greenhouse-Geisser statistic indicated significant differences, then the Games-Howell post-hoc analysis was conducted to determine where the differences occurred. When significant

differences were found within trials, paired, post-hoc *t*-tests were conducted to determine where the differences occurred, with alpha set at .01.

Results

Demographic information. Table 1 presents the demographic characteristics for the sample. Group differences for gender and highest grade completed were analyzed using a Chi-square test for independence. There were no significant differences found for gender, X^2 (2, n = 62) = .21, p = .90. Cramer's V = .06, or for highest grade completed, X^2 (4, n = 62) = 4.37, p = .36. Cramer's V = .19. Group differences for age were analyzed using one-way ANOVA. There were no significant differences found for age, F(2, 59) = .27, p = .77 (eta squared = .01). Group differences for race/ethnicity could not be analyzed because there were not a sufficient number of participants in the non-White groups.

Table 1

Demographic Characteristics of Study Sample

	Group				
	Coached	Uncoached	Controls	X^2 or F	p
Gender				.21	.90
Males (*n*)	4	5	4		
Females (*n*)	18	16	15		
Education				4.37	.36
High School	8	7	3		
Some College	12	14	14		
College Degree	2	0	2		
Age					
M (*SD*)	20.59 (*3.40*)	20.57 (*3.27*)	19.95 (*2.64*)	2.7	.77
Race/Ethnicity					
European American (*n*)	20	20	17		
Native American (*n*)	1	0	1		
Asian American (*n*)	1	0	1		
Two or more races (*n*)	0	1	0		

Primary Analysis

PVMT. A two-way ANOVA was conducted with group and trial as independent variables. Mauchley's Test of Sphericity indicated that the assumption of sphericity had been violated, thus a Greenhouse-Geisser correction was used to interpret the F statistic. There was no significant interaction effect between group and trial, $F(2, 59) = 2.15$, $p = .105$. There was a statistically significant main effect for group, $F(2, 59) = 18.50$, $p = .000$, partial eta squared = .39 (large effect). The assumption of equal variances was not met, therefore the Games-Howell post-hoc analysis was conducted. The Games-Howell analysis revealed that mean scores for both the coached and uncoached simulators were significantly lower than mean scores for the control group on all PVMT trials. Mean scores for the coached and uncoached simulators were not significantly different from each other on all PVMT trials. The means and standard deviations for each group on each trial of the PVMT are presented in Table 2 and Figure 1.

Table 2

Mean Number Correct on Each Trial of the PVMT for Each Group

	Group				
	Controls *M*(*SD*)	Coached *M*(*SD*)	Uncoached *M*(*SD*)	*F*	*p*
PVMT					
Trial 1	43.37(1.17)$_a$	32.95(8.36)$_b$	27.05(10.73)$_b$	18.50	.000
Trial 2	44.63(0.68)$_a$	30.68(11.03)$_b$	26.48(14.08)$_b$		
Trial 3	43.84(1.04)$_a$	29.05(11.62)$_b$	25.05(14.70)$_b$		

Note. Means in the same row having the same subscript are not significantly different at $p < .01$ in the post-hoc comparisons.

Figure 1 shows the mean number correct on each trial of the PVMT for each group.

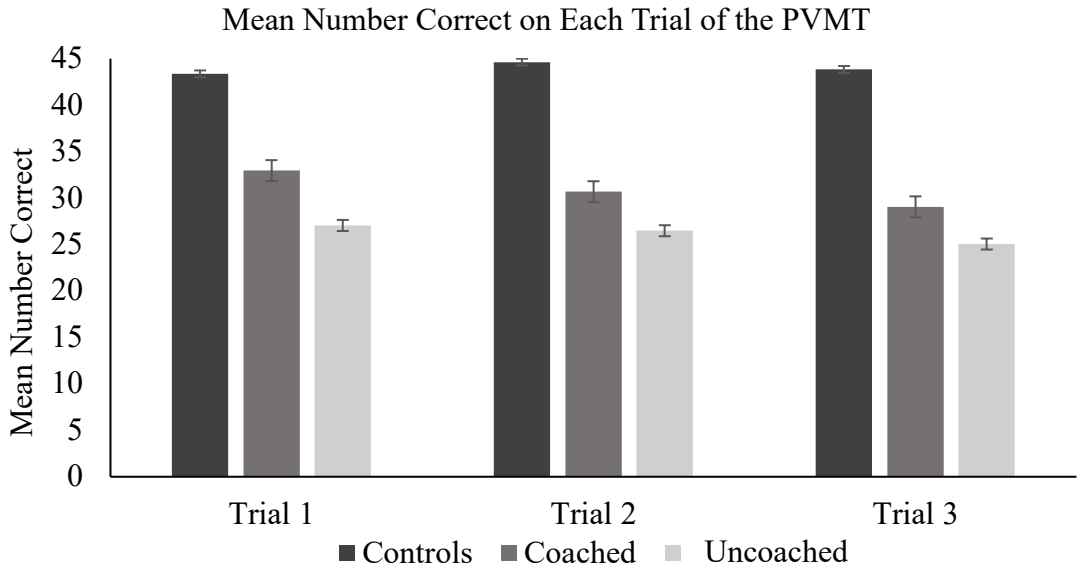


Figure 1. Mean Number Correct on Each Trial of the PVMT

The Greenhouse-Geisser correction was again used to interpret the F statistic for the effect of trial. There was a statistically significant main effect for trial, $F(1, 59) = 4.18$, $p = .031$, partial eta squared = .07 (medium effect). Post-hoc comparisons were made with three, paired samples *t*-tests to identify which trials were significantly different from one another. There was a statistically significant decrease in PVMT scores from Trial 2 ($M = 33.53$, $SD = 12.87$) to Trial 3 ($M = 32.23$, $SD = 13.46$), $t(61) = 2.90$, $p < .01$ (two-tailed). The mean decrease in PVMT scores was 1.31 with a 95% confidence interval ranging from .41 to 2.21. The calculated eta squared statistic (.12) indicated an intermediate effect (Cohen, 1988). The means and standard deviations for each pair of trials are presented in Table 3.

Table 3

Paired samples T-tests

		M(SD)	*t*	*p*
Pair 1	Trial 1 Correct	34.15(10.32)$_a$	.986	.328
	Trial 2 Correct	33.53(12.87)$_a$		
Pair 2	Trial 2 Correct	33.53(12.87)$_a$	2.90	**.005**
	Trial 3 Correct	32.23(13.46)$_b$		
Pair 3	Trial 1 Correct	34.15(10.32)$_a$	2.29	.025
	Trial 3 Correct	32.23(13.46)$_a$		

Note. Means in the same pair having the same subscript are not significantly different at $p < .01$ in the post-hoc comparisons.

Exploratory Analyses

One-way, between groups ANOVA was conducted to investigate PVMT completion times for each group and to examine group performance across other tests in the battery. For exploratory analyses, if the variances were greater than 6 times each other, the Welch test was used to interpret the F ratio. If the Welch Test revealed significant differences between the groups, Games-Howell post-hoc analysis was conducted to determine where the differences occurred. If the Welch Test was not necessary, no correction was applied, and the Tukey HSD test was conducted for post-hoc comparisons. For all exploratory analyses, alpha was set at .05.

PVMT Trial 1: Time to Complete. Levene's statistic for one-way ANOVA with Trial 1 (total time in seconds as the dependent variable) revealed that the assumption of homogeneity was not met. As a result, the Welch Test was used to examine differences. The Welch Test showed a statistically significant difference at the $p < .05$ level between the groups, F (2, 58) = 17.82, $p = .000$. The effect size, calculated using eta squared, was 0.29 (large effect). Post-hoc

comparisons using the Games-Howell test indicated that mean times for the coached group (M = 195.10, SD = 73.26) and the uncoached group (M = 180.86, SD = 53.06) were significantly greater than the mean time for the control group (M = 113.85, SD = 32.80). Mean times for the coached group and the uncoached group were not significantly different from one another. The means and standard deviations for total response time for Trial 1 can be found in Table 4.

PVMT Trial 2: Time to Complete. Levene's statistic for one-way ANOVA with Trial 2 (total time in seconds as the dependent variable) revealed that the assumption of homogeneity was not met. As a result, the Welch Test was again used to examine differences. The Welch Test showed a statistically significant difference at the $p < .05$ level between the groups, F (2, 58) = 18.98, $p = .000$. The effect size, calculated using eta squared, was 0.27 (large effect). Post-hoc comparisons using the Games-Howell test indicated that mean times for the coached group (M = 197.95, SD = 85.45) and the uncoached group (M = 181.48, SD = 65.65) were significantly greater than the mean time for the control group (M = 108.38, SD = 25.79). Mean times for the coached group and the uncoached group were not significantly different from one another. The means and standard deviations for total response time for Trial 1 can be found in Table 4.

PVMT Trial 3: Time to Complete. Levene's statistic for the one-way ANOVA with Trial 3 (total time in seconds as the dependent variable) revealed that the assumption of homogeneity was not met. As a result, the Welch Test was again used to examine differences. The Welch Test showed a statistically significant difference at the $p < .05$ level between the groups, F (2, 57) = 18.44, $p = .000$. The effect size, calculated using eta squared, was 0.26 (large effect). Post-hoc comparisons using the Games-Howell test indicated that mean times for the coached group (M = 207.52, SD = 84.32) and the uncoached group (M = 195.35, SD = 83.99) were significantly greater than the mean time for the control group (M = 113.54, SD = 25.44). Mean times for the

coached group and the uncoached group were not significantly different from one another. The means and standard deviations for total response time for Trial 1 can be found in Table 4.

Table 4

Mean Time (in seconds) on PVMT Trials

	Group				
	Controls *M*(*SD*)	Coached *M*(*SD*)	Uncoached *M*(*SD*)	*F*	*p*
PVMT					
Trial 1	113.85(32.80)$_a$	195.10(73.26)$_b$	180.86(53.06)$_b$	17.82	.000
Trial 2	108.38(25.79)$_a$	197.95(85.45)$_b$	181.48(65.65)$_b$	18.98	.000
Trial 3	113.54(25.44)$_a$	207.52(84.32)$_b$	195.35(83.99)$_b$	18.44	.000

Note. Means in the same row having the same subscript are not significantly different at $p < .05$ in the post-hoc comparisons.

Trail Making Test A. Levene's statistic for one-way, between-groups ANOVA with Trails A completion time (in seconds) as the dependent variable revealed that the assumption of homogeneity was not met. As a result, the Welch test was used to examine differences between the groups. The Welch Test showed a statistically significant difference at the $p < .05$ level between the groups, F (2, 59) = 11.90, p = .000. The effect size, calculated using eta squared, was 0.18; (large effect). Post-hoc comparisons using the Games-Howell test indicated that mean time scores for the coached group (M = 50.82, SD = 27.53) and the uncoached group (M = 52.05, SD = 30.62) were significantly greater than the mean time score for the control group (M = 26.35, SD = 10.97). Mean time scores for the coached group and the uncoached group were not significantly different from one another. The means and standard deviations for Trails A time are presented in Table 5.

Trail Making Test B. Levene's statistic for one-way, between-groups ANOVA with Trails B completion time (in seconds) as the dependent variable revealed that the assumption of homogeneity was not violated. Results showed a statistically significant difference at the $p < .05$ level between the groups, F (2, 59) = 4.20, p = .020. The effect size, calculated using eta squared, was 0.12; (medium effect). Post-hoc comparisons using the Tukey HSD test indicated that mean time scores for the coached group (M = 105.41, SD = 71.53) and the uncoached group (M = 108.19, SD = 63.49) were significantly greater than the mean time score for the control group (M = 60.63, SD = 22.49). Mean scores for the coached group and the uncoached group were not significantly different from one another. The means and standard deviations for Trails B time are presented in Table 5.

Table 5

Mean Time on Trail Making Tests A and B

	Group				
	Controls *M*(*SD*)	Coached *M*(*SD*)	Uncoached *M*(*SD*)	*F*	*p*
TMT					
Time A (s)	$26.35(10.97)_a$	$50.82(27.54)_b$	$52.05(30.62)_b$	11.90	.000
Time B (s)	$60.63(22.49)_a$	$105.41(71.53)_b$	$108.19(63.49)_b$	4.20	.020

Note. Means in the same row having the same subscript are not significantly different at $p < .05$ in the post-hoc analyses.

KBNA-Cancellation Subtest: Time to Complete. Levene's statistic for one-way, between-groups ANOVA with Cancellation completion time (in seconds) as the dependent variable revealed that the assumption of homogeneity was not met. As a result, the Welch Test

was used to examine differences between the groups. The Welch Test showed a statistically significant difference at the $p < .05$ level between the groups, F (2, 59) = 11.51, p = .000. The effect size, calculated using eta squared, was 0.16 (large effect). Post-hoc comparisons using the Games-Howell test indicated that mean time scores for the coached group (M = 98.23, SD = 51.29) and the uncoached group (M = 99.76, SD = 66.71) were significantly greater than the mean time score for the control group (M = 51.68, SD = 18.49). Mean time scores for the coached group and the uncoached group were not significantly different from one another. The means and standard deviations for Cancellation time are presented in Table 6.

KBNA- Cancellation Subtest: Total Correct. Levene's statistic for one-way, between-groups ANOVA with Cancellation total correct as the dependent variable revealed that the assumption of homogeneity was not violated. There were no statistically significant differences in the total correct score at the $p < .05$ level between the groups, F (2, 59) = 2.33, p = .11. The means and standard deviations for Cancellation total correct are presented in Table 6.

Table 6

Mean Times and Total Correct on KBNA-Cancellation

	Group				
	Controls *M*(*SD*)	Coached *M*(*SD*)	Uncoached *M*(*SD*)	*F*	*p*
Cancellation					
Time (s)	$51.68(18.49)_a$	$98.23(51.29)_b$	$99.76(66.71)_b$	11.51	.000
Total Correct	$58.63(1.64)_a$	$58.50(2.24)_a$	$57.14(3.18)_a$	2.33	.11

Note. Means in the same row having the same subscript are not significantly different at $p < .05$ in the post-hoc analyses.

WRAT4- Word Reading Subtest. Levene's statistic for one-way, between-groups ANOVA with Word Reading total correct as the dependent variable revealed that the assumption of homogeneity was not violated. There were no statistically significant differences in the total correct score at the $p < .05$ level between the groups, F (2, 59) = .72, p = .49. The means and standard deviations for Word Reading total correct are presented in Table 7.

Table 7

Mean Number Correct on WRAT4- Word Reading Subtest

	Group				
	Controls *M*(*SD*)	Coached *M*(*SD*)	Uncoached *M*(*SD*)	*F*	*p*
Word Reading Total Correct	$61.26(4.34)_a$	$59.32(5.32)_a$	$60.24(5.48)_a$	.72	.50

Note. Means in the same row having the same subscript are not significantly different at $p < .01$ in the post-hoc analyses.

Aphasia Screening Test. Levene's statistic for one-way, between-groups ANOVA with AST total correct as the dependent variable revealed that the assumption of homogeneity was not met. As a result, the Welch Test was used to examine differences between the groups. The Welch test showed a statistically significant difference at the $p < .05$ level between the groups, F (2, 59) = 12.01, p = .000. The effect size, calculated using eta squared, was 0.18 (large effect). Post-hoc comparisons using the Games-Howell test indicated that mean correct scores for the coached group (M = 28.00, SD = 2.83) and the uncoached group (M = 27.33, SD = 3.98) were significantly less than the mean correct score for the control group (M = 30.47, SD = 1.02). Mean correct scores for the coached group and the uncoached group were not significantly

different from one another. The means and standard deviations for AST total correct are presented in Table 8.

Table 8

Mean Number Correct on Aphasia Screening Test

	Group				
	Controls *M*(*SD*)	Coached *M*(*SD*)	Uncoached *M*(*SD*)	*F*	*p*
AST					
Total Correct	$30.47(1.02)_a$	$28.00(2.83)_b$	$27.33(3.98)_b$	12.01	.000

Note. Means in the same row having the same subscript are not significantly different at $p < .05$ in the post-hoc analyses.

TOMM Trial 1. Levene's statistic for one-way, between-groups ANOVA with TOMM Trial 1 total correct as the dependent variable revealed that the assumption of homogeneity was not met. As a result, the Welch Test was used to examine differences between the groups. The Welch Test showed a statistically significant difference at the $p < .05$ level between the groups, $F(2, 59) = 38.81$, $p = .000$. The effect size, calculated using eta squared, was 0.43 (large effect). Post-hoc comparisons using the Games-Howell test indicated that mean correct scores for the coached group (M = 35.09, SD = 8.82) and the uncoached group (M = 29.86, SD = 10.93) were significantly less than the mean correct score for the control group (M = 47.42, SD = 3.13). Mean time scores for the coached group and the uncoached group were not significantly different from one another. The means and standard deviations for TOMM Trial 1 total correct are presented in Table 9.

TOMM Trial 2. Levene's statistic for one-way, between-groups ANOVA with TOMM Trial 2 total correct as the dependent variable revealed that the assumption of homogeneity was not met. As a result, the Welch test was used to examine differences between the groups. The Welch test showed a statistically significant difference at the $p < .05$ level between the groups, $F(2, 59) = 34.95$, $p = .000$. The effect size, calculated using eta squared, was 0.37 (large effect). Post-hoc comparisons using the Games-Howell test indicated that mean correct scores for the coached group (M = 36.59, SD = 10.95) and the uncoached group (M = 30.62, SD = 13.81) were significantly less than the mean correct score for the control group (M = 49.79, SD = .92). Mean correct scores for the coached group and the uncoached group were not significantly different from one another. The means and standard deviations for TOMM Trial 2 total correct are presented in Table 9.

Table 9

Mean Number Correct on TOMM Trial 1 and TOMM Trial 2

	Group				
	Controls *M*(*SD*)	Coached *M*(*SD*)	Uncoached *M*(*SD*)	*F*	*p*
Trial 1 Total Correct	$47.42(3.13)_a$	$35.09(8.82)_b$	$29.86(10.93)_b$	38.81	.000
Trial 2 Total Correct	$49.79(.92)_a$	$36.59(10.94)_b$	$30.62(13.81)_b$	34.95	.000

Note. Means in the same row having the same subscript are not significantly different at $p < .05$ in the post-hoc analyses.

Post-test Questionnaire 1 (PTQ1). Questions 1-5 of PTQ1 asked participants to rate their performance across testing using 5-point, Likert scales. These questions included whether they followed instructions, how hard they tried to follow instructions, how successful they thought

they were in following instructions, whether they thought they were successful in keeping the examiner from knowing their group, and how familiar they were with the effects of mTBI before participation. The results of questions 1-5 are presented below in figures 2-6.

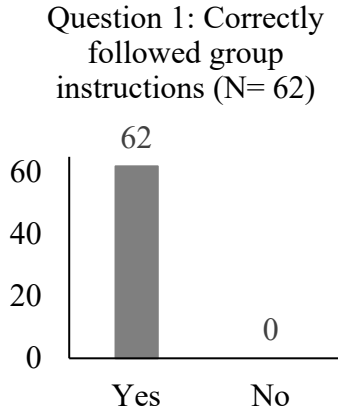


Figure 2. Number of participants that correctly followed instructions

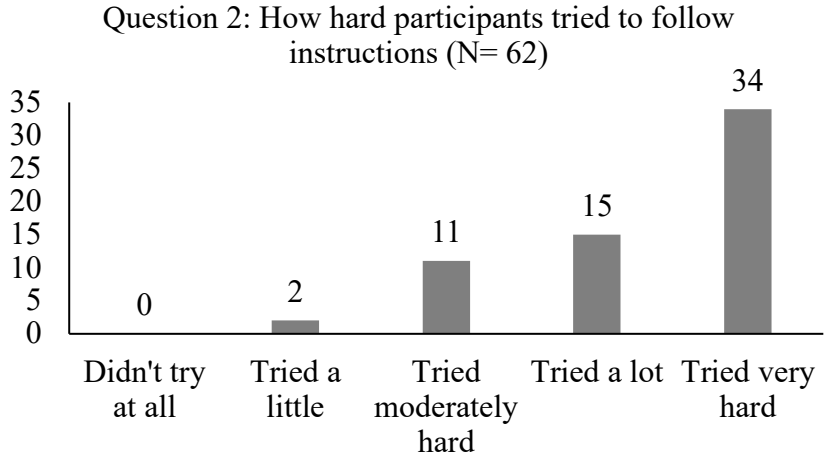


Figure 3. How hard participants tried to follow instructions

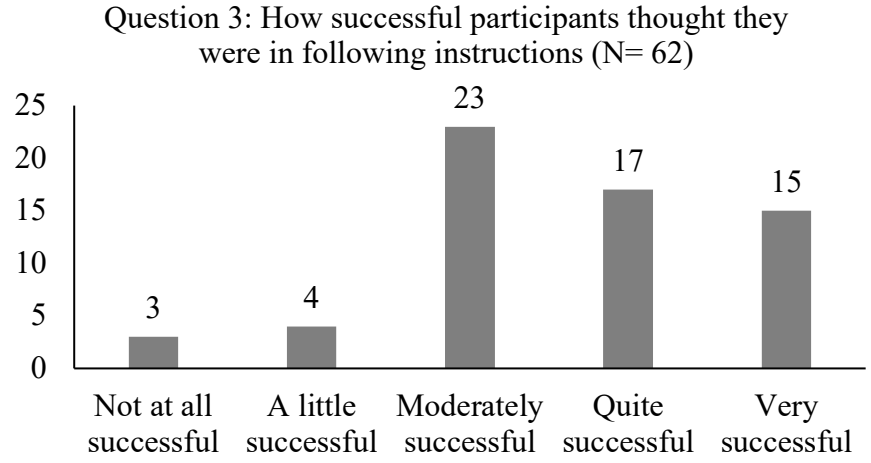


Figure 4. How successful participants thought they were in following instructions

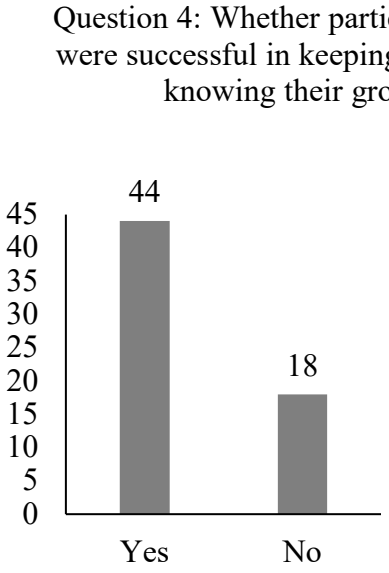


Figure 5. Whether participants thought they were successful in keeping the examiner from knowing their group

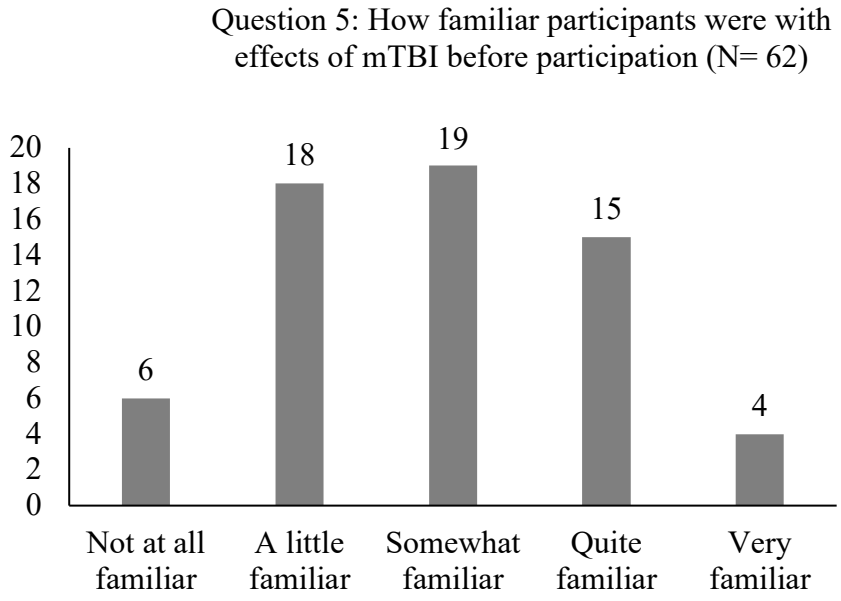


Figure 6. How familiar participants were with the effects of mild TBI before participation

Questions 6-12 of PTQ1 asked participants to identify one main construct they thought each test measured. Two researchers coded responses and inter-rater reliability was assessed using a model put forth by McAlister, Lee, Ehlert, Kajfez, Faber, and Kennedy (2017). As recommended by the model, a code book was generated with numerous, predetermined, well-defined codes. Those codes were: memory; speed; attention/concentration; speech/language; motor; planning, organizing, and problem-solving; comprehension, and other. The researchers used the code book to code all qualitative data for questions 6-12 of PTQ2. The code book with codes can be found in Appendix K. Once the data were coded, inter-rater reliability (IRR) was calculated for each test following the formula described in Miles and Huberman (1994) that says that an IRR of 80% agreement between coders on 95% of the total codes is sufficient agreement among multiple coders. Results gave 86% agreement on 95% of the total codes. IRR for each test can be found in Appendix L. The data are presented below in table 10.

Table 10

PTQ1:Questions 6-12: Purpose of each test

	Tests in the Battery													
Codes	**TOMM**		**Word Reading**		**TMT A**		**TMT B**		**AST**		**Cancellation**		**PVMT**	
	A*	B*	A	B	A	B	A	B	A	B	A	B	A	B
Memory	61	61	8	8	8	9	9	9	24	22	23	19	62	62
Speed	0	0	1	0	22	18	14	14	3	0	19	17	0	0
Attention/Concentration	0	0	1	2	7	8	12	14	3	4	11	9	0	0
Speech/Language	0	0	33	27	1	0	4	0	3	4	0	0	0	0
Motor	0	0	0	1	3	3	1	1	3	1	0	1	0	0
Problem Solving	0	1	0	8	8	14	11	17	2	9	0	5	0	0
Comprehension	0	0	7	10	4	0	2	1	5	4	2	1	0	0
Other	1	0	12	6	9	10	9	6	19	18	6	10	0	0
TOTAL	62	62	62	62	62	62	62	62	62	62	62	62	62	62

A = Rater 1, B = Rater 2

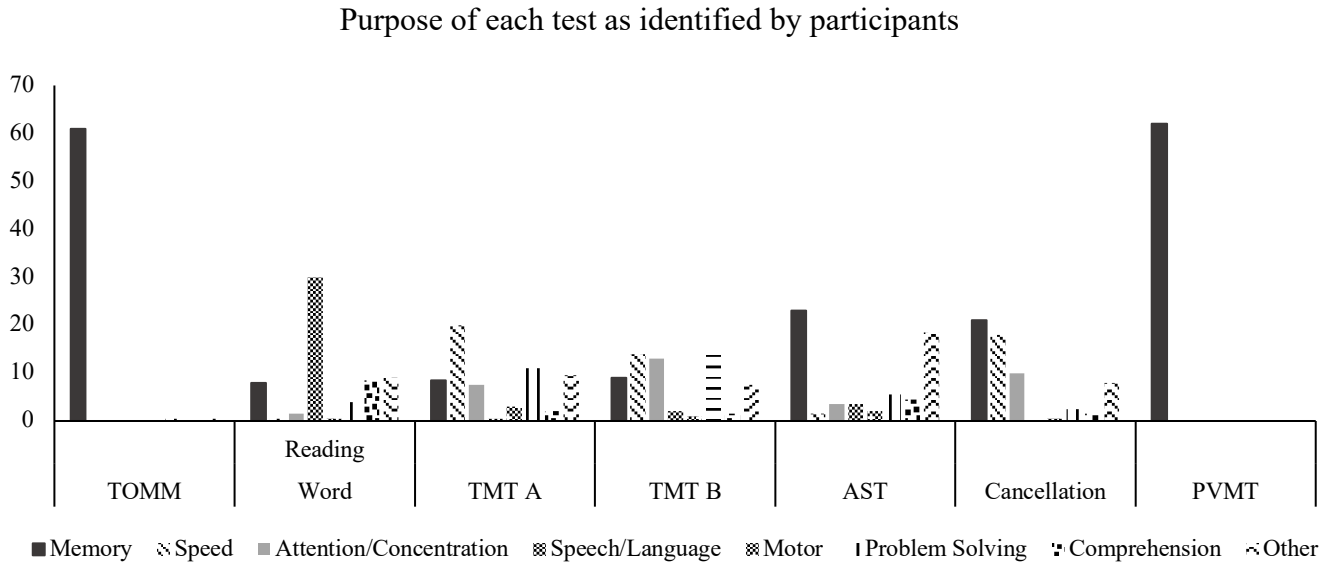


Figure 7. Purpose of each test as identified by participants

Post-test Questionnaire 2. PTQ2 stated that some of the tests in the battery were designed to detect individuals faking brain damage, while other tests are typically administered to assess cognitive abilities such as attention, memory, and processing speed. Participants were asked to place a check mark by any test that seemed as if it were designed to detect someone faking brain damage. If participants marked any tests, they were asked to indicate how certain they were that the marked test was designed to detect faked brain damage. Results are presented in Table 11, Figure 8, Table 12, and Figure 9.

Table 11

PTQ2: Number of times each test was identified as a test of effort

Response	TOMM	Word Reading	TMT A	TMT B	AST	Cancellation	PVMT
No	23	18	21	26	14	20	23
Yes	38	43	40	35	47	41	38
Total*	61	61	61	61	61	61	61

*One participant did not complete this question (N = 61)

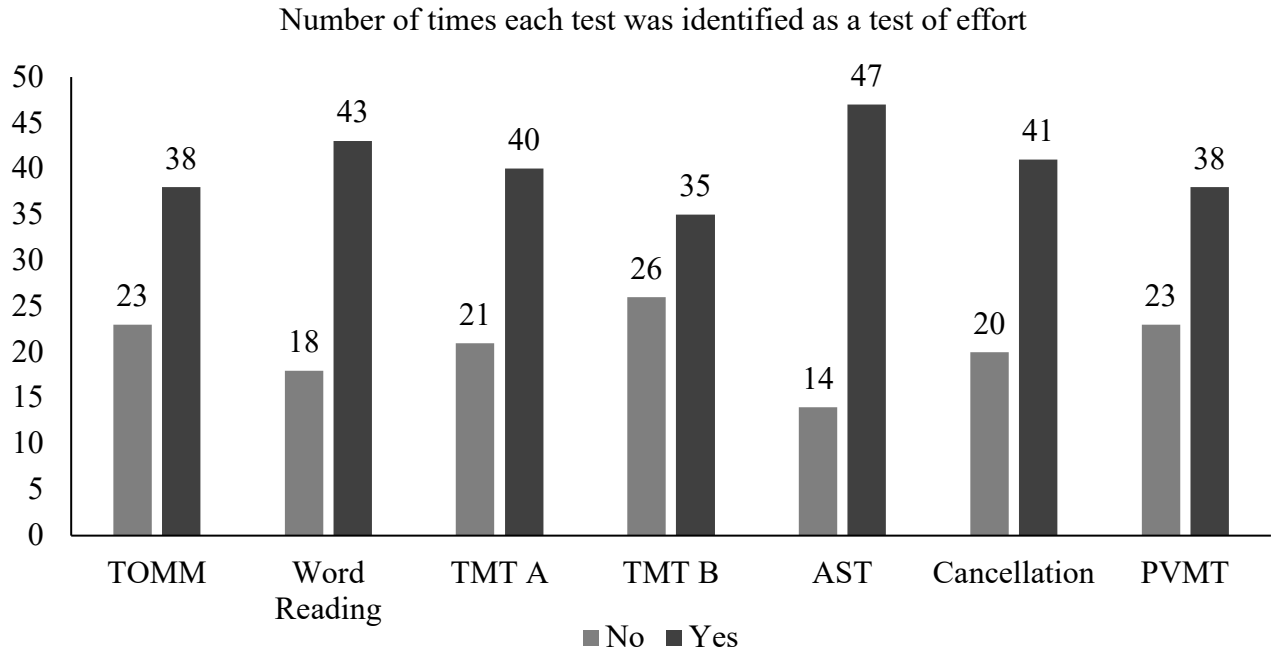


Figure 8. Number of times each test was identified as a test of effort

Table 12

PTQ2: Participants' certainty in their judgement(s) of which tests measured effort

Test	**1 Not at all certain**	**2 A little certain**	**3 Somewhat certain**	**4 Fairly certain**	**5 Very Certain**	**Total times identified as effort test**
TOMM	1	10	8	13	6	38
Word Reading	0	4	13	19	7	43
TMT A	2	6	18	8	6	40
TMT B	0	7	11	13	4	35
AST	0	8	19	17	3	47
Cancellation	2	8	22	5	4	41
PVMT	2	6	12	12	6	38

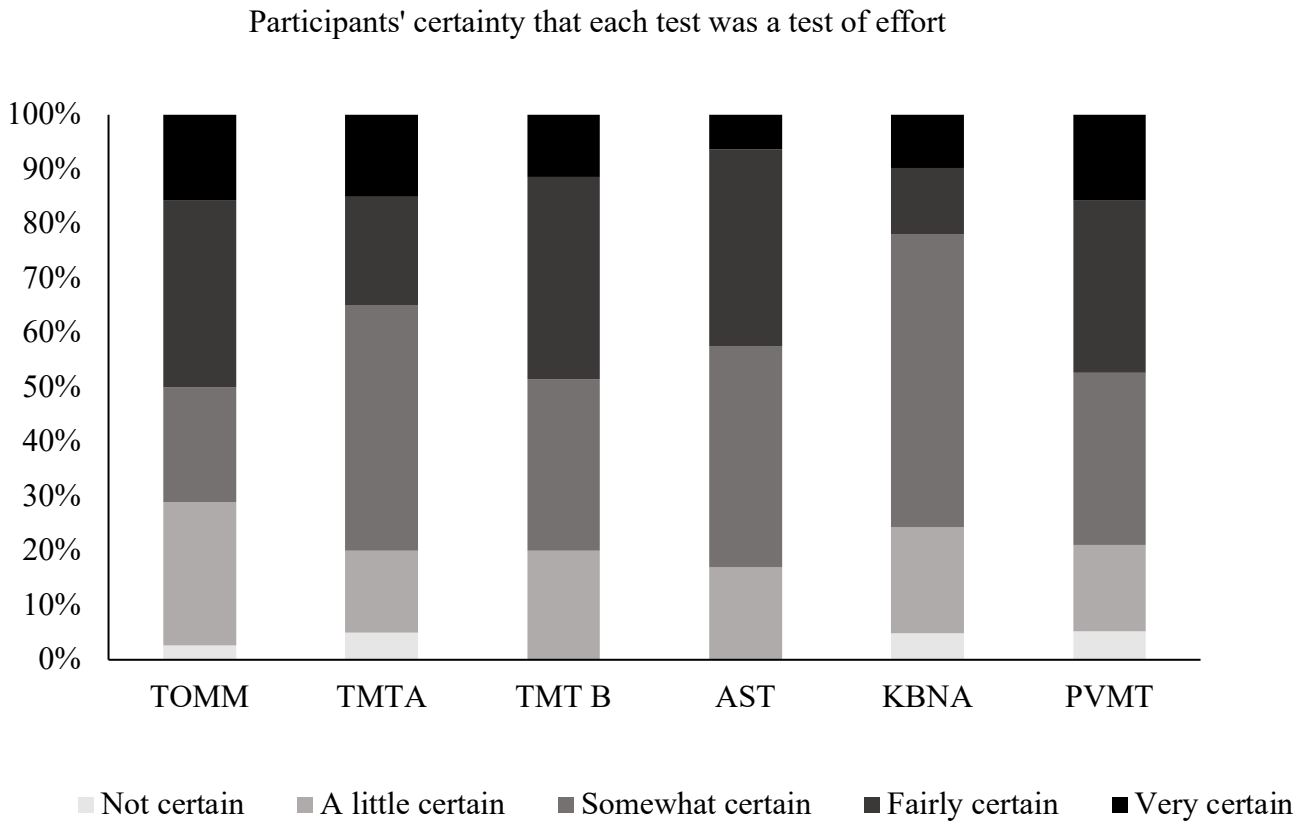


Figure 9. Participants' certainty that each test was a test of effort

Receiver Operating Characteristic (ROC) Analyses. A ROC curve was generated for each trial of the PVMT (3 trials) to identify sensitivity and specificity at various cut scores and to provide a scale for classifying the accuracy of the test via the area under the curve (AUC). An AUC value closer to 1 would indicate the test reliably distinguished between true positives and false positives, whereas an AUC value closer to .50 would indicate the test was no more accurate than chance (Zhou, Obuchowski, & Obuchowski, 2002). For the PVMT, the AUC for each trial was as follows: trial 1 was .94, trial 2 was .92, and trial 3 was .91. The following guide (Mandrekar, 2010) was used for classifying the AUC of each trial: .90-1 was considered 'excellent,' .80-.89 was considered 'good,' .70-.79 was considered 'fair,' .60-.69 was considered

'poor,' and .50-.59 was considered 'fail.' AUC values for each trial of the PVMT were 'excellent' and indicated that the test accurately discriminated simulators and non-simulators. Next, coordinates of each ROC curve were examined to determine acceptable cut scores for each trial. Cut scores on each trial that provided a sensitivity of .90 and a specificity of at least .70 were selected according to recommendations offered by Boone (2013) and Larrabee (2012). Based on this contingency, the cut score for trial 1 was 43/45, the cut score for trial 2 was 44/45, and the cut score for trial 3 was 43/45. ROC curves for each trial of the PVMT, AUC for each curve, and a complete table of possible cut scores for each trial of the PVMT may be found in Appendix M.

This same process was repeated with the TOMM whereby ROC curves were generated for each trial, and the AUC for each trial was plotted. For the TOMM, the AUC for trial 1 was .92 and the AUC for trial 2 was .92, both 'excellent' (Mandrekar, 2010). The cut score for trial 1 of the TOMM was 44/50 and the cut score for trial 2 of the TOMM was 49/50. ROC curves for each trial of the TOMM, AUC for each curve, and a complete table of possible cut scores for each trial of the TOMM may be found in Appendix N.

Discussion

The primary goals of this project were to create a new test of performance validity (the PVMT), to validate the PVMT with neurologically normal individuals, and to examine the performance characteristics of the PVMT with control subjects and individuals simulating cognitive impairment. Additional goals were to explore the sensitivity and specificity of the PVMT at various cut scores and to compare the accuracy of the PVMT with a well-validated PVT (the TOMM) by examining area under the curve (AUC) for each test.

Most importantly, neurologically normal individuals putting forth their best effort (control subjects), successfully completed each trial of the PVMT with minimal errors. This finding confirms that the test operated optimally, and as expected. This finding also clearly demonstrated that the novel design features of the PVMT did not negatively impact control subjects' performance.

The first novel design feature included a shift away from the conventional two-alternative forced choice designs of many commonly used PVTs, to a progressive three- and four-alternative forced choice design. This design feature moves away from reliance on below chance responding (found in two-alternative designs) as the primary method for identifying dissimulation. Research has shown that most malingerers do not perform so poorly that they score statistically significantly below chance levels (Binder, 2002; Loring, Larrabee, Lee, & Meador, 2007). As a consequence of using a below chance criterion, malingerers' scores may be misinterpreted as representing their best efforts, when in fact dissimulation occurred. To avoid this problem, PVT researchers have begun to use empirically derived cut scores based on memory-impaired clinical populations whose scores on PVTs have been shown to be very similar to healthy controls (Green, 2003; Tombaugh, 1996). Therefore, dissimulators scoring above chance may still be accurately identified as dissimulators if they also score below the cut score set by memory-impaired groups. Incorporation of empirically derived cut scores is a design feature found in the PVMT that serves to maximize the test's sensitivity and specificity, while also promoting increased diversity in the types of research designs found in PVT test development.

The second novel design feature involved the number of foils per item and the number of trials. It was anticipated that increasing the number of foils within each item and the total number of trials would increase the face validity of the PVMT as a test of memory, without actually

making increased demands on participants' memory. Results showed that these additions did not negatively impact control subjects' performance; in other words, control subjects were able to produce near perfect scores on all trials of the PVMT.

The third novel design feature involved the actual PVMT stimuli. All stimuli within each item were semantically related (e.g., the target and foils were all birds, or, the target and foils were all mailboxes, etc.). It was assumed that semantic similarity between stimuli on each item of the PVMT would further increase the perception of the PVMT as a genuine test of memory without impacting performance. Again, control subjects were able to produce near perfect scores on all trials of the PVMT.

Hypothesis 1 predicted that mean PVMT scores for the coached simulators would be significantly lower than scores for the control group on each trial of the PVMT. Results revealed that Hypothesis 1 was supported. This finding is interesting because it demonstrates that even when provided with additional information about how to more accurately portray cognitive impairment, coached simulators were nevertheless readily detectible. This finding is somewhat encouraging given research that has shown that coaching can result in more sophisticated and difficult to detect malingers (Suhr & Gunstad, 2000; Vickery, Berry, Dearth, Vagnini, & Baser, 2004). Two possible explanations for this finding are offered. First, it may be that the scope of the information provided about how to portray mTBI was not sufficient, therefore coached simulators were not able to put it to use. This inference is drawn from a meta-analysis conducted by Suhr and Gunstad (2007) that found that the most successful coached malingerers were coached with multiple types of information including information about brain injury symptoms as well as information about the tests themselves. In this case, only information about brain injury symptoms was offered. Second, it is possible that the coached simulators were not able to

quickly translate the instructions into a nuanced, sophisticated performance without practice (Brennan, Meyer, David, Pella, Hill, & Gouvier, 2009).

Hypothesis 2 predicted that mean PVMT scores for the uncoached simulators would be significantly lower than scores for the control group on each trial of the PVMT. Results revealed that Hypothesis 2 was supported. This result was not surprising given that uncoached simulators did not receive additional information about how to subtly portray brain injury and instead relied upon their own knowledge of mTBI during testing. Many participants in this group performed more like individuals with severe TBI or a severe neurodegenerative condition, an effect that has been cited in prior simulation studies (Binder, 1993; Hiscock & Hiscock, 1989; Lezak et al., 2012; Rees, Tombaugh, Gansler, & Moczynski, 1998; Russeler, Brett, Klaue, Sailer, & Munte, 2008).

Hypothesis 3 predicted that mean PVMT scores for the uncoached simulators would be significantly lower than scores for the coached simulators on each trial of the PVMT. Results revealed that Hypothesis 3 was not supported. Intriguingly, there were no significant differences found between uncoached and coached simulators on any test administered. This fact reveals more about the performance of the coached simulators than the uncoached simulators and provides further support for the idea that coached simulators were not able to use additional information to their advantage.

Hypothesis 4A predicted that mean PVMT scores for uncoached simulators would decrease at a significantly greater rate than the coached simulators across the three trials while hypothesis 4B predicted coached simulators will decrease at a significantly greater rate than the control group across the three trials. Results revealed there were no differential effects across trials for any of the individual groups, thus these hypotheses were not supported. Graphs of the

data demonstrate no significant change for any group in mean scores from trial 1 to trial 2 and from trial 2 to trial 3. This finding may be related to the instructions read aloud to participants in between each trial of the PVMT. Instructions signaled increasing difficulty with each trial, though in actuality difficulty remained equivalent across trials as evidenced by the performance of the control group. For example, instructions stated, "This trial may be more difficult than the previous trial," and, "This trial may be the most difficult trial." This test design feature can also be found in the TOMM, a well-established and highly used PVT. However, simulators' performance was not influenced by this strategy. Perhaps the language used was not sufficiently persuasive in this case, or perhaps the visually resonant nature of the stimuli overrode these subtle cues to perform poorly. Lastly, it may be that this type of cue is not as effective outside of real-world situations in which poor performance is highly financially incentivized and has been shown to increase with increasing financial incentives (Bianchini, Curtis, & Greve, 2006; Larrabee, 2012).

In contrast to the finding that there were no differential effects across trials for any of the individual groups, there *was* a significant effect due to trial when the groups were combined. Paired samples *t*-tests revealed that this difference occurred between trials 2 and 3. This was surprising as visual inspection of the data do not suggest a significant difference in the means for trials 2 and 3. Furthermore, the mean difference between trial 1 and trial 3 was larger than the mean difference between trial 2 and trial 3, yet the mean difference between trial 1 and trial 3 was not found to be significant. Therefore, this finding may be anomalous, and interpretation is reserved until further studies replicate this finding.

ROC Analyses. Other analyses produced additional results that support the quality of the PVMT. Primary among these were receiver operating characteristic (ROC) analyses that

examined sensitivity and specificity at all possible cut points on each trial of the PVMT, identified the area under the curve for each trial of the PVMT (as indices of overall test accuracy), and compared these with those found on the TOMM. Cut scores for each trial of the PVMT were selected that provided a sensitivity of at least 0.90 and a specificity of at least 0.70 (Boone, 2013; Larrabee, 2012). Area under the curve (AUC) for each trial of the PVMT was found to be at least 0.90, indicating excellent ability to distinguish true positives from false positives (Mandrekar, 2010). These findings are extremely similar to those found for the TOMM on this project (see Appendices N and O) which suggests that the two tests are functioning similarly. This is compelling given that the TOMM is considered an excellent PVT for malingering and is widely used in neuropsychological evaluations (Rees, Tombaugh, Gansler, & Moczynski, 1998; Vallabhajosula & van Gorp, 2001).

Post-test Questionnaires. Additional interesting findings were derived from post-test questionnaires. Post-test questionnaire 1 asked participants to identify the purpose of each test. Participants uniformly identified the PVMT as a test of memory. This finding supports the strategy described earlier in which attempts were made to increase the face validity of the PVMT as a test of memory with the inclusion of more foils per item and an additional 3rd trial. To date, no published studies were found that expressly investigated face validity in this fashion.

Post-test questionnaire 2 stated that certain tests in the battery measured effort and asked participants to identify, which tests, if any, were effort tests. Results demonstrated that even when given this information, participants were no more likely to identify the PVMT as an effort test than any other test in the battery. In fact, participants rated all the tests remarkably similarly with roughly twice as many endorsements (that a test was an effort test) as denials (that a test was not an effort test) for each test. In other words, participants were not able to make accurate

determinations about which tests measured effort, and which did not. This is promising as it shows participants were not able to identify the PVMT as a test of effort.

PVMT Completion Times. Total completion times for each group for each trial of the PVMT were investigated and results revealed a pattern similar to that seen when analyses were conducted on the total number correct for each trial. Mean scores for the coached and uncoached groups were significantly different from the control group (simulators were slower than the controls and easily identifiable), and the coached and uncoached groups were not significantly different from each other. The fact that simulators were easily distinguished from controls based on time to complete, points to the utility in capturing these data. Prior research has shown that response time can be used effectively as an indicator of sub-optimal test performance (Kim, Boone, Victor, Marion, & Amano, 2010; Rose, Hall, Szalda-Petree, & Bach, 1998; Willison, & Tombaugh, 2006). The results also revealed that the coached simulators unexpectedly performed more slowly than the uncoached simulators. Confusion about how to incorporate the extra information given to them may have produced slower times, or perhaps they preferred to go slow over making errors as a way to demonstrate impairment. Research has shown that combining response latency with number of errors could further enhance the sensitivity and specificity of the PVMT (Rose, Hall, Szalda-Petree, & Bach, 1998).

Other Test Data. Significant differences were found between simulators and the controls on four of the five tests given (Trail Making, KBNA, AST, and TOMM), and findings may provide insight into the strategy used by simulators to simulate brain injury. When a test was timed, participants worked extremely slowly as opposed to making numerous errors. When a test was not timed, participants evidenced impairment by making many errors. While this seems intuitive, it underscores an important finding; that slowing down may be the most accessible and

least complicated method for simulating brain injury (e.g., when in doubt about what to do, go slow). As mentioned above, this is consistent with literature that has determined response time to be an indicator of non-effortful performance (Kim, Boone, Victor, Marion, & Amano, 2010; Rose et al., 1998; Willison, & Tombaugh, 2006). One test, WRAT4-Word Reading, did not follow the above pattern of either slowing down or increasing errors. In this untimed test, participants read aloud increasingly difficult vocabulary words. Vocabulary is considered a form of crystallized knowledge and is highly correlated with intellectual ability where higher vocabulary scores suggest greater intellectual abilities. There were no significant differences between the groups on Word Reading which suggests that the simulators did not make any attempts to appear impaired. This is interesting from a PVT development standpoint. One possible explanation may be that it is more difficult for individuals to simulate brain injury effectively on tests that tap crystallized, long-term knowledge, such as vocabulary items; however, this is an empirical question that requires study.

In addition to the findings listed above, other aspects of the PVMT were demonstrated to be advantageous. For example, benefits were observed for the test administrator. First, computerization likely made the PVMT easier to administer than the TOMM which is a pencil and paper test and requires that the tester turn the pages of a booklet at a specific rate of speed. While page-turning is not unduly difficult, using an automated program removes any error that might occur due to fluctuations in human timing. Computerization also unburdens the test administrator during recognition trials in which the administrator must turn pages, record responses, and provide verbal feedback on each response. Lastly, computerization precludes the need to replace worn or lost materials such as the three spiral books of stimuli.

Limitations

This project aimed to validate the PVMT with neurologically normal individuals putting forth their best effort. While the project was successful in this regard, there are large challenges yet to overcome. Foremost among these is the need to administer the PVMT with important clinical groups such as those with demonstrated memory impairment as research has shown that individuals from these groups are able to perform comparably to healthy controls on performance validity tests (Allen, Iverson, & Green, 2002; Rees, Tombaugh, Gansler, & Moczynski, 1998; Tombaugh, 1996). Without such a group, there would not be normative data to which test-takers' performance could be compared. Next steps are underway to begin administration of the PVMT with a clinical group with mixed memory impairments. Data from clinical groups would also facilitate a more rigorous comparison with the TOMM, as the TOMM is normed on a clinical sample (Tombaugh, 1996).

A second limitation involves the use of simulation design. It has been argued that simulator performance is not generalizable to the performance of individuals deliberately feigning impairment because simulators and dissimulators are not equally motivated and do not have the same outcome expectations (Rogers, 2008; Ruiz, Drake, Glass, Marcotte, & van Gorp, 2002). Steps were taken to make the incentive offered to participants as meaningful as possible. In the current study, participants were offered additional research credits if they were able to portray mTBI in such a way as to remain undetected as a simulator. This incentive is obviously not comparable to the large financial settlements found in "real-world" situations; however, attempting to make incentives equivalent is not possible. Ultimately, it may be more helpful to simply keep in mind the limitations of simulation design while striving to make simulation experiments as externally valid as possible.

A third limitation involved the fact that the test battery did not include a genuine test of memory functioning. Had there been one, it might have affected participants' judgements about the purpose of each test and their judgements about which tests were tests of effort. For example, perhaps more participants would have identified the PVMT and TOMM as tests of effort, given their experience taking an actual memory test. In other words, the face validity of the PVMT and TOMM as tests of memory might be more variable than that observed in this experiment.

Conclusion

The detection of exaggerated or malingered symptoms is a complex issue that neuropsychologists regularly encounter during evaluations across a variety of contexts (Bianchini, Mathias, & Greve, 2001; Drob, Meehan, & Waxman, 2009; Greiffenstein, Baker, & Gola, 1994; Mittenberg, Patton, Canyock, & Condit, 2002; Pella et al., 2012; Rees, Tombaugh, Gansler, & Moczynski, 1998; Teichner, & Wagner, 2004). It is therefore critical, that neuropsychologists utilize instruments that are capable of detecting invalid or false claims of impairment if an accurate diagnosis and appropriate treatment recommendations are to be made.

The purpose of this project was to create a new, stand-alone PVT designed to detect malingered memory performance, to validate the test with healthy controls, and to examine the performance characteristics of the test with controls and two groups simulating cognitive impairment. Findings demonstrated that the PVMT performed largely as expected, with healthy controls easily completing each trial successfully, and simulators easily detected. Additional findings provided insight into subjects' perceptions of the tests in the battery, including their judgments about which tests were intended to measure effort. In general, participants were not able to accurately identify the true underlying purpose of each test, nor could they accurately identify which tests were actual tests of effort. This finding may alleviate some concern among

clinicians about the extent of individuals' knowledge regarding commonly administered neuropsychological tests, their ability to recognize tests of effort, and their ability to successfully dissimulate. Other findings were also encouraging. For example, ROC analyses demonstrated high sensitivity and specificity on each trial of the PVMT and suggest that it may perform similarly to the TOMM, a widely used test of performance validity (Rees, Tombaugh, Gansler, & Moczynski, 1998; Vallabhajosula & van Gorp, 2001). Whether or not the PVMT performs as expected with important clinical groups remains unknown; however, experimental results show promise. Continued efforts in performance validity test development will help to preserve the ability of neuropsychologists to make accurate statements about cognitive functioning and will ensure that patients are recommended appropriate treatments.

References

Allen, L. M., Iverson, G. L., & Green, P. (2002). Computerized Assessment of Response Bias in forensic neuropsychology. *Journal of Forensic Neuropsychology, 3*(1/2), 205-225.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Arlington, VA, American Psychiatric Association, 2013.

Arnold, G., & Boone, K. B. (2007). Use of motor and sensory tests as measures of effort. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neurological perspective* (pp. 178-209). New York, NY: Guilford Press.

Ashendorf, L., Vanderslice-Barr, J. L., & McCaffrey, R. J. (2009). Motor tests and cognition in healthy older adults. *Applied Neuropsychology, 16,* 171-176.

Babin, P. R., & Gross, P. (2002). When symptoms don't add up: Conversion and malingering in rehabilitation settings. *Journal of Rehabilitation, 68,* 4-13.

Balddessarini, R. J., Finkelstein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40*(5), 569-573.

Barrash, J., Suhr, J., & Manzel, K. (2004). Detecting poor effort and malingering with an expanded version of the Auditory Verbal Learning Test (AVLTX): Validation with clinical samples. *Journal of Clinical and experimental Neuropsychology, 26*(1), 125-140.

Benton, A. L., Hamsher, K. de S., & Sivan, A. B. (1994). Multilingual Aphasia Examination. Iowa City: AJA Associates.

Bianchini, K. J., Curtis, K. L., & Greve, K. W. (2006). Compensation and malingering in traumatic brain injury: A dose-response relationship? *Clinical Neuropsychology, 20,* 831-847.

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist, 15*(1), 19-45.

Bigler, E. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society, 18,* 632-642.

Bigler, E. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury, 28,* 1623-1638.

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 15,* 170-183.

Binder, L. M. (2002). The Portland Digit Recognition Test: A review of validation data and clinical use. *Journal of Forensic Neuropsychology*, *2*, 27-41.

Binder, L. M., & Pankratz, L. (1987). Neuropsychological evidence of a facticious memory complaint. *Journal of Clinical and Experimental Neuropsychology, 9,* 167-171.

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. *The American Journal of Psychiatry, 153,* 7-10.

Boone, K. (2013). *Clinical practice of forensic neuropsychology*. New, NY: Guilford.

Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razini, J. (2002). The Rey 15-Item recognition trial: A technique to enhance sensitivity of the Rey 15-Item Memorization Test. *Journal of Clinical and Experimental Neuropsychology, 24,* 561-573.

Brennan A. M., Meyer, S., David, E., Pella, R., Hill, B. D., & Gouvier, W. D. (2009). The vulnerability to coaching across measures of effort. *The Clinical Neuropsychologist, 23*, 314-328.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity, NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20,* 419-426.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., et al. (1989, revised 2001). *Minnesota Multiphasic Personality Inventory*. San Antonio, TX: Pearson.

Centers for Disease Control and Prevention, 2003. Report to Congress on Mild Traumatic Brain Injury in the United States: Steps to Prevent a Serious Public Health Problem. Atlanta, GA: National Center for Injury Prevention and Control.

Cicerone, K. D., & Azulay, J. (2002). Diagnostic utility of attention measures in post-concussion syndrome. *The Clinical Neuropsychologist, 16*(3), 280-289.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 284-7.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test – Second Edition (CVLT-II)*. San Antonio, TX: Psychological Corporation.

Dikman, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of the expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*, 5, 346-356.

Donders, J., & Strong, C. A. (2011). Embedded effort indicators on the California Verbal Learning Test – Second Edition (CVLT-II): An attempted cross-validation. *The Clinical Neuropsychologist, 25,* 173-184.

Drob, S. L., Meehan, K. B., & Waxman, S. E. (2009). Clinical and Conceptual Problems in the Attribution of Malingering in Forensic Evaluations. *The Journal of the American Academy of Psychiatry and the Law, 37:* 98-106.

Dwyer, C. A. (1996). Cut scores and testing: Statistics, judgement, truth, and error. *Psychological Assessment, 8,* 360-362.

Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: an amerging paradox of embedded validity indicators. *The Clinical Neuropsychologist, 31*: 6-7, 1029-1046.

Erdodi, L. A., Seke, K. R., Shahein, A., Tyson, B. T., Sagar, S., & Roth, R. M. (2017). Low scores on the Grooved Pegboard Test are associated with invalid responding and psychiatric symptoms. *Psychology & Neuroscience, 10*(3), 325-344.

Essig, S. M., Mittenberg, W., Petersen, R. S., Strauman, S., & Cooper, J. T. (2001). Practices in forensic neuropsychology: Perspectives of neuropsychologists and trial attorneys. *Archives of Clinical Neuropsychology, 16,* 271-191.

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

Faust, D., & Ackley, M. A. (1998). Did you think it was going to be easy? Some methodological suggestion for the investigation and development of malingering detection techniques. In C. R. Reynolds (Ed.), *Detection of malingering during head injury litigation* (pp. 261-286). New York: Plenum Press.

Ferguson, K. E. (2003). Detecting malingering in forensic neuropsychological evaluations in litigants with mild traumatic brain injury. In W. T. O'Donohue and E. R. Levensky (Eds.), *Handbook of forensic psychology*. London, UK: Elsevier Academic Press.

Franzen, M. D., & Iverson, G. L. Detecting negative response bias and diagnosing malingering. In P. J. Snyder and P. D. Nussbaum (Eds.), *Clinical Neuropsychology: A Pocket Handbook for Assessment.* Washington, DC: American Psychological Association; 1998.

Frederick, R. A. (2002). A review of Rey's strategies for detecting malingered neuropsychological impairment. *Journal of Forensic Neuropsychology, 2,* 1-25.

Glosser, G., Butters, N., & Kaplan, E. (1977). Visuoperceptual processes in brain-damaged patients on the Digit Symbol Substitution Test. *International Journal of Neuroscience*, 7, 59-66.

Gouvier, W. D. (1999). Base rates and clinical decision making in neuropsychology. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice* (pp. 27-37). Lisse, The Netherlands: Swets & Zeitlinger.

Gouvier, W. D., Lees-Haley, P., & Hayes, J. H. (2003). The neuropsychological examination in the problem of detecting malingering in the forensic arena: Costs and benefits. In Prigatano & N. Pliskin (Eds.), *Clinical neuropsychology and cost outcomes research: A beginning* (pp. 405-424). New York, NY: The Psychology Press.

Green, P. (2003). *Manual for the Word Memory Test.* Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2013). Spoiled for choice: Making comparisons between forced-choice effort tests. In K. B. Boone (Ed.), *Clinical practice of forensic neuropsychology*. New York, NY: Guilford Press.

Green, P., Lees-Haley, P., & Allen, L. M. (2002). The Word Memory Test in the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology, 2,* 97-124.

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

Greve, K. W., & Bianchini, K. J. (2004). Setting empirical cut-offs on psychometric indicators of negative response bias: A methodological commentary with recommendations. *Archives of Clinical Neuropsychology, 19*(4), 533-541.

Greve, K. W., Bianchini, K. J., Mathias, C. W., Hourson, R. J., & Crouch, J. A. (2003). Detecting malingered performance on the Wechsler Adult Intelligence Scale: Validation of Mittenberg's approach in traumatic brain injury. *Archives of Clinical Neuropsychology, 18*(3), 245-260.

Hartlage, L. C. Clinical Detection of Malingering. In C. R. Reynolds and A. M., Horton, Jr. (Eds.), Detection of Malingering during Head Injury Litigation. Springer Science and Business Media, LLC, 2012.

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Gerrard, M. O. M. (2007). Retesting in selection: A meta-analysis of coaching and practicing effects for tests of cognitive ability. *The Journal of Applied Psychology, 92*(2), 373-385.

Heaton, R. K., Ryan, L., & Grant, I. (2009). Demographic influences and use of demographically corrected norms in neuropsychological assessment. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric and neuromedical disorders (3rd ed.).* New York: Oxford University Press.

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

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. The Clinical Neuropsychologist, 23, 1093-1129.

Hennekens, C. H., & Buring, J. E. (1987). *Epidemiology in medicine*. Boston: Little, Brown, and Company.

Henry, J. D., & Beatty, W. W. (2006). Verbal fluency deficits in Multiple Sclerosis. *Neuropsychologia, 44*, 1166-1174.

Henry, J. D., & Crawford, J. R. (2004b). Verbal fluency deficits in Parkinson's disease: A meta-analysis. *Journal of the International Neuropsychological Society, 10*, 608-622.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004a). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia, 42*, 1212-1222.

Hepp, U., Gamma, A., Milos, G., Eich, D., Ajdacic-Gross, V., Rossler, W., et al. (2006). Inconsistency in reporting potentially traumatic events. *The British Journal of Psychiatry, 188,* 278-283.

Hinton-Bayre, A. D., Geffen, G., & McFarland, K. (1997). Mild head injury and speed of information processing: A prospective study of professional rugby league players. *Journal of Clinical and Experimental Neuropsychology*, 19, 275-289.

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11*(6), 967-974.

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont, California: Cengage Wadsworth.

Husky, M. (2002). The Portland Digit Recognition Test, The Test of Memory Malingering, and Standard Neuropsychological Measures: Comparing efficacy, the effects of coaching, and face validity. *Graduate student master's thesis.*

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

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

Iverson, G. L. & Franzen, M. D. (1996). Using multiple objective measures to detect simulated malingering. Journal of Clinical and Experimental Neuropsychology, 18, 38–51.

Kim, M. S., Boone, K. B., Victor, T., Marion, S. D., & Amano, S. (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. *Archives of Clinical Neuropsychology, 25*(1), 60-70.

Kimpton, T. (2014). Simulation of Brain Injury: Preparation outside the lab. *Graduate Student Theses, Dissertations, & Professional Papers.* http://scholarworks.umt.edu/etd/4578.

Langeluddecke, P. M., & Lucas, S. K. (2003). Quantitative measures of memory malingering on the Wechsler Memory Scale- Third Edition in mild head injury litigants. *Archives of Clinical Neuropsychology, 18,* 181-197.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 22,* 1-14.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18,* 625-631.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. *Neuropsychological Assessment – Fifth Edition.* Oxford University Press: New York, NY, 2012.

Loring, D. W., Larrabee, G. J., Lee, G. P., & Meador, K. J. (2007). Victoria Symptom Validity Test performance in a heterogenous clinical sample. *The Clinical Neuropsychologist, 21*(3), 522-531.

Lynch, W. J. (2004). Determination of effort level, exaggeration, and malingering in neurocognitive assessment. *Journal of Head Trauma Rehabilitation, 19*(3), 277-283.

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American Professionals. *The Clinical Neuropsychologist, 29*:6, 741-776.

Martin, R. C., Hayes, J. S., & Gouvier, W. D. (1996). Differential vulnerability between postconcussion self-report and objective malingering tests in identifying simulated mild head injury. *Journal of Clinical and Experimental Neuropsychology, 18*, 265-275.

Matthews, C. G., & Klove, K. (1964). *Instruction manual for the Adult Neuropsychology Test Battery.* Madison, WI: University of Wisconsin Medical School.

McAlister, A. M., Lee, D. M., Ehlert, K. E., Kajfez, R. L., Faber, C. J., & Kennedy, M. S. (2017). Qualitative Coding: An approach to assess inter-rater reliability. *American Society for Engineering Education,* Conference Paper.

McCaffrey, R. J., Lynch, J. K., & Howe, L. L. S. Explaining Symptom Validity Testing to the Trier of Fact. In C. R. Reynolds and A. M. Horton (Eds.), *Detection of Malingering during Head Injury Litigation.* Springer Science and Business Media, LLC, 2012.

Miles, M. B., & Huberman, A. M. *Qualitative data analysis: An expanded sourcebook – Second Edition.* Los Angeles: Sage Publications, 1994.

Millon, T. (1977). *Millon Clinical Multiaxial Inventory*. Minneapolis, MN: National Computer Systems.

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

Morey, L. C. *Personality Assessment Inventory* Professional Manual. Odessa, FL: Psychological Assessment Resources: 1991.

Nies, K. J., & Sweet, J. L. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology, 9*(6), 501-552.

Orey, S. A., Cragar, D. E., & Berry, D. T. R. (2000). The effects of two motivational manipulations on the neuropsychological performance of mildly head-injured college students. *Archives of Clinical Neuropsychology, 15,* 335-348.

Pallant, J. *SPSS Survival Manual 5th ed.* New York, NY: McGraw-Hill: 2013.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology, 47*(2), 409-410.

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory impairment. *Perceptual and Motor Skills, 57,* 367-372.

Pankratz, L., & Binder, L.M. (1997). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception, Second Edition* (pp. 223-236). New York: The Guillford Press.

Pella, R. D., Hill, B.D., Ashvind, N. S., Hayes, J. S., & Gouvier, W. D. (2012) Noncredible Performance in Mild Traumatic Brain Injury. In C. R. Reynolds and A. M. Morton, Jr. (Eds.), Detection of Malingering during Head Injury Litigation, Springer Science and Business Media, LLC.

Ponsford, J., Camerson, P, Fitzgerald, M., Grant, M., Mikocka-Walus, A., & Schonberger, M. (2012). Predictors of Postconcussive Symptoms 3 Months After Mild Traumatic Brain Injury. *Neuropsychology, 26*(3), 304-313.

Prigatano, G. P. (1999). *Principles of neuropsychological rehabilitation*. Oxford, UK: Oxford University Press.

Promberger, M., & Marteau, T. M. (2013). When do financial incentives reduce intrinsic motivation? Comparing behaviors studied in psychological and economic literatures. *Health Psychology: Official Journal of the Division of Health Psychology*, American Psychological Association, *32*, 950-957.

Raskin, S. A., & Rearick, E. (1996). Verbal fluency in individuals with mild traumatic brain injury. *Neuropsychology, 10*, 416-422.

Rees, L. M., Tombaugh, T. N., & Boulay, L. (2001). Depression and the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 16*, 501-506.

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,* 10-20.

Reitan, R. M., & Wolfson, D. (1993). The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation (2nd ed.). Tucson, AZ: Neuropsychology Press.

Resnick, P. J. (1997). The detection of malingered mental illness. *Behavioral Science and Law, 2,* 20-38.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives of Psychology, 28, 286-340.*

Reynolds, C. R. (1998). Common sense, clinicians, and actuarialism in the detection of malingering during head injury litigation. In C. R. Reynolds (Ed.), *Detection of malingering during head injury litigation* (pp. 261-286). New York: Plenum.

Reynolds, C. R., & Horton, A. M. Clinical Acumen, Common Sense, and Data-Based Decision Making in the Assessment of Dissimulation During Head Injury Litigation. In C. R. Reynolds and A. M. Horton (Eds.), *Detection of Malingering during Head Injury Litigation.* Springer Science and Business Media, LLC, 2012.

Reynolds, M. K., Clark, B. E., & Hall, S. (2017). A significant threat to neuropsychological test validity. Unpublished manuscript.

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

Rogers, R. (Ed.). (2008). *Clinical assessment of malingering and deception* (3rd ed.). New York: Guilford Press.

Root, J. C., Robbins, R. N., Chang, L., & Van Gorp, W. G. (2006). Detection of inadequate effort on the California Verbal Learning Test- Second Edition: Forced Choice Recognition and critical item analysis. *Journal of the International Neuropsychological Society, 12*, 688-696.

Rose, F. E., Hall, S., & Szalda-Petree, A. (1998). A comparison of four tests of malingering and the effects of coaching. *Archives of Clinical Neuropsychology, 13*(4), 349-363.

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.

Russeler, J., Brett, A., Klaue, U., Sailer, M., & Munte, T. (2008). The effect of coaching on the simulated malingering of memory impairment. *BMC Neurology, 8*(37).

Schwartz, E. S., Erdodi, L., Rodriguez, N., Ghosh, J. J., Curtain, J. R., Flashman, L. A., & Roth, R. M. (2016). CVLT-II forced choice recognition trial as an embedded validity indicator: A systematic review of the evidence. *Journal of the International Neuropsychological Society, 22,* 851-858.

Silver, J. M. (2015). Invalid symptom reporting and performance: What are we missing? *Neurorehabilitation, 36,* 463-469.

Slick, M. K., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical research and practice. *The Clinical Neuropsychologist, 13,* 45-561.

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

Sohlberg, M. M., & Mateer, C. A. (2001). *Cognitive rehabilitation: An integrative neuropsychological approach.* New York, NY: Guilford.

Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology, 26,* 744-789.

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

Suhr, J., & Gunstad, J. (2007). Coaching and malingering: A review. In G. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (287-311). New York, NY: Oxford University Press.

Taylor CA, Bell JM, Breiding MJ, Xu L. (2017). Traumatic Brain Injury–Related Emergency Department Visits, Hospitalizations, and Deaths — United States, 2007 and 2013. MMWR Surveillence Summaries, 66(No. SS-9):1–16.

Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): normative data from cognitively intact, cognitively impairmed, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, 19, 455-464.

Tombaugh, T. N. (1995). *Test of memory malingering* (TOMM). Toronto, ON: Multi-Health Systems.

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

Vallabhajosula, B., & van Gorp, W. G. (2001). Post Daubert admissibility of scientific evidence on malingering of cognitive deficits. *Journal of the American Academy of Psychiatry and the Law, 29*, 207-215.

Vickery, C. D., Berry, D. T. R., Dearth, C. S., Vagnini, V. L., & Baser, R. E. (2004). Head injury and the ability to feign neuropsychological deficits. *Journal of Clinical and Experimental Neuropsychology, 15,* 578-590.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology, 16,* 45-73.

Warrington, E. K. (1984). *Recognition memory Test*. Los Angeles: Western Psychological Services.

Wechsler, D. *Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV).* Pearson: 2008.

Wechsler, D. *Wechsler Memory Scale – Fourth Edition (WMS-IV)*. Pearson: 2009.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test 4 (WRAT4)*. Lutz, FL: Psychological Assessment Resources.

Williams, J. M., & Jones, K. Factitious Responding and Malingered Memory Disorder. In C. R. Reynolds and A. M. Horton (Eds.), *Detection of Malingering during Head Injury Litigation.* Springer Science and Business Media, LLC, 2012.

Willis, P. F., Farrer, T.J., & Bigler, E. D. Are effort measures sensitive to cognitive impairment? *Military Medicine, 176,* 1426-1431.

Willison, J. & Tombaugh, T. N. (2006). Detecting simulation of attention deficits using reaction time tests. *Archives of Clinical Neuropsychology, 21*(1), 41-52.

Yates, A.J. (1954). The validity of some psychological tests of brain damage. *Psychological Bulletin, 51, 359-379.*

Youngjohn, J. R. (1995). Confirmed attorney coaching prior to neuropsychological evaluation. *Assessment, 2*(3), 279-383.

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

Zhou, X. H., Obuchowski, N. A., & Obushcowski, D. M. (2002). *Statistical methods in diagnostic medicine.* Wiley & Sons: New York.

Appendix A: Informed Consent Form

Study Title: Progressive Visual Memory Test (PVMT)

Investigator(s):

Brook Clark, M.A., Principal Investigator
brook.clark@umconnect.umt.edu
406-243-4521

Stuart Hall, Ph.D., Faculty Advisor
stuart.hall@umconnect.umt.edu
406-243-4521

Special Instructions:

This consent form may contain words that are new to you. If you read any words that are not clear to you, please ask the person who gave you this form to explain them to you.

Inclusion Criteria:

You must be 18 or older to participate in this research.

Purpose:

You are being asked to take part in a research study that examines the test characteristics of various neuropsychological tests. The results will be used facilitate the development of a new neuropsychological test.

Procedures:

If you agree to take part in this research study, informed consent will be reviewed, and you will be given an envelope containing instructions for how to complete the rest of the study. You will then complete tests of thinking and problem-solving *according to the instructions you received in the envelope*. After all testing, you will complete two Post-Test Questionnaires. The session will last a maximum of 60 minutes and will take place in Skaggs Building room 246.

Compensation for Participation:

You will receive **4 research credits** for participation in this study.

Risks/Discomforts:

There is no anticipated discomfort for those contributing to this study, so risk to participants is minimal. However, answering questions on the demographic questionnaire may make you feel sad or upset. Therefore, a list of resources in community will be provided to you at the end of participation.

Benefits:

Although you may not directly benefit from taking part in this study, results may benefit scientific knowledge.

Confidentiality:

Your records will be kept confidential and will not be released without your consent except as required by law. If the results of this study are written in a scientific journal or presented at a scientific meeting, your name will not be used. All data will be stored in a locked file cabinet.

Your signed Informed Consent Form will be stored in a cabinet separate from the data to ensure anonymity.

Voluntary Participation/Withdrawal:

Your decision to take part in this research study is entirely voluntary.
You may leave the study for any reason without penalty.
You may be asked to leave the study for any of the following reasons:

1. Failure to follow the Project Director's instructions;
2. A serious adverse reaction which may require evaluation;
3. The Project Director thinks it is in the best interest of your health and welfare; or
4. The study is terminated.

Questions:

If you have any questions about the research now or during the study, please contact Stuart Hall, Ph.D. at 406-243-4521. If you have any questions regarding your rights as a research subject, you may contact the UM Institutional Review Board (IRB) at (406) 243-6672.

Statement of Your Consent:

I have read the above description of this research study. I have been informed of the risks and benefits involved, and all my questions have been answered to my satisfaction. Furthermore, I have been assured that any future questions I may have will also be answered by a member of the research team. I voluntarily agree to take part in this study. I understand I will receive a copy of this consent form.

_____
Printed Name of Subject

_____ _____
Subject's Signature Date

Disclosure of Personal Health Information:

I authorize Brook Clark, M.A., and the researcher's staff, to use the individual health information collected on my demographic questionnaire, for the purpose of conducting this research project.

_____
Printed Name of Subject

_____ _____
Subject's Signature Date

Appendix B: Demographic Questionnaire

***INSTRUCTIONS**: Please complete the following screening questionnaire by filling in the blanks or circling your answers.*

Age: _____ Birthdate: _____ Gender: _____ Ethnicity: _____

1. Were there any known difficulties with your birth? Yes No

   If yes, describe: _____

2. Do you have a vision problem that requires corrective lens wear (e.g., glasses)? Yes No

Education

3. Did you ever have to repeat any grades? Yes No

4. Were you ever placed in special education classes? Yes No

5. What is the highest level of education you have attained (circle one)?

   High School Some College College degree Master's degree Doctoral degree

Medical and Health History

6. Have you ever been diagnosed with any neurological condition? Yes No

   If yes, or unsure, please list: _____

7. Are you currently experiencing significant problems with your mental health, such as problems with anxiety and/or depression, or any other psychiatric condition? Yes No
   If yes, please list:

   _____

   _____

   _____

8. Are you currently receiving treatment for your mental health? Yes No
   If yes, please explain:

_____

_____

_____

9. Have you ever felt you should cut down on your drinking/drug use? Yes No

10. Have you ever been annoyed by people who criticize your drinking/drug use? Yes No

11. Have you ever felt bad or guilty about your drinking or drug use? Yes No

12. Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hang over? Yes No

13. Do you often drive under the influence of alcohol or drugs? Yes No

14. Have you ever been diagnosed with cancer? Yes No

15. If yes, what type of cancer did you have? _____

16. If yes, when were you diagnosed with cancer? _____

17. If yes, how long did your cancer treatment last? _____

18. If yes, what kind of cancer treatment did you receive?

_____

_____

_____

19. If yes, are you finished will all treatments for this cancer? Yes No

20. If yes, when did you finish all cancer treatments? _____

21. Have you ever experienced a concussion or brain injury? Yes No

22. Were you knocked unconscious? Yes No

If yes, how long were you unconscious? (please circle the letter that corresponds to your answer)

A. Less than 1 minute

B. 1-30 minutes

C. More than 30 minutes

23. Do you remember all of the events **before** your head injury? Yes No

If **no**, how much time passed between your last memory and when the injury occurred?

1. A few seconds
2. Less than 5 minutes
3. Less than 30 minutes
4. 30 to 60 minutes
5. More than 60 minutes

24. Do you remember all of the events **after** your head injury? Yes No

If **no**, how much time passed between the injury and the next thing you remember?

6. A few seconds
7. Less than 5 minutes
8. Less than 30 minutes
9. 30 to 60 minutes
10. More than 60 minutes

Thank you.

Appendix C: Instructions for Uncoached Simulators

Please follow these instructions for the remainder of the experiment **without letting the researcher know what you have been told to do**. When you finish reading these instructions, please sign at the bottom indicating that you have read the instructions and understand what to do. Then, place this signed page back into the envelope, seal it, place an X over the seal, and wait for the examiner to return.

Please read the following instructions carefully:

You are about to take some cognitive tests that examine mental abilities such as attention, concentration, problem-solving, and your ability to think quickly. While you are taking the tests, please pretend that you have experienced a concussion from a car accident where you were rear-ended while stopped at a traffic light. The force of the collision caused you to bump your head on the steering wheel, but you did not lose consciousness. You noticed you felt "dazed," but you were able to speak and follow conversations normally. However, you found it difficult to recall the details of the accident and the events that occurred immediately after it. You were transported to the hospital where you were diagnosed with a concussion. Symptoms of concussion include headache, fatigue, and trouble thinking. You were discharged with instructions to rest and avoid watching television or using your computer for several days. While resting at home, you decide to go to court to get money from the driver who was responsible for hitting your car. Over the next few months, the negative symptoms from your concussion end, however your lawsuit has not yet been settled. Your attorney has mentioned that you may win more money if it appears that you are still suffering from the concussion.

As you pretend to be this car accident victim, try to respond to each test as if you were trying to appear brain injured in order to get money from a lawsuit. Therefore, your performance on the tests should convince the researcher, and officials deciding the outcome of your lawsuit, that you are still suffering from the concussion. **If you are able to convince the researcher that you have a brain injury, you will receive two additional research credits, for a total of 6 credits.**

Please sign your name at the bottom of this page, place it back into the envelope, seal the envelope, and place an X over the seal before the researcher returns. **Remember, do not let the researcher know what you have been told to do. Your performance on the tests should be the only way of figuring it out.**

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

_____

Signature and Date

Appendix D: Instructions for Coached Simulators

Please follow these instructions for the remainder of the experiment **without letting the researcher know what you have been told to do**. When you finish reading these instructions, please sign at the bottom indicating that you have read the instructions and understand what to do. Then, place this signed page back into the envelope, seal it, place an X over the seal, and wait for the examiner to return.

Please read the following instructions carefully:

You are about to take some cognitive tests that examine mental abilities such as attention, concentration, problem-solving, and your ability to think quickly. While you are taking the tests, please pretend that you have experienced a concussion from a car accident where you were rear-ended while stopped at a traffic light. The force of the collision caused you to bump your head on the steering wheel, but you did not lose consciousness. You noticed you felt "dazed," but you were able to speak and follow conversations normally. However, you found it difficult to recall the details of the accident and the events that occurred immediately after it. You were transported to the hospital where you were diagnosed with a concussion. Symptoms of concussion include headache, fatigue, and trouble thinking. You were discharged with instructions to rest and avoid watching television or using your computer for several days. While resting at home, you decide to go to court to get money from the driver who was responsible for hitting your car. Over the next few months, the negative symptoms from your concussion end, however your lawsuit has not yet been settled. Your attorney has mentioned that you may win more money if it appears that you are still suffering from the concussion.

As you pretend to be this car accident victim, try to respond to each test as if you were trying to appear brain injured in order to get money from a lawsuit. Therefore, your performance on the tests should convince the researcher, and officials deciding the outcome of your lawsuit, that you are still suffering from the concussion. **If you are able to convince the researcher that you have a brain injury, you will receive two additional research credits, for a total of 6 credits.**

Here is a helpful hint for how to be successful in this task:

Try to simulate the most severe problems you can *without making it too obvious to the researcher!* If you appear too impaired, such as not being able to remember things or think at all, the researcher will easily detect your efforts. Your goal is to convince the researcher that you are still suffering from the concussion, because this information will be used to obtain money in court.

Please sign your name at the bottom of this page, place it back into the envelope, seal the envelope, and place an X over the seal before the researcher returns. **Remember, do not let the researcher know what you have been told to do. Your performance on the tests should be the only way of figuring it out.**

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

_____

Signature and Date

Appendix E: Instructions for the Control Group

Please follow these instructions for the remainder of the experiment **without letting the researcher know what you have been told to do**. When you finish reading these instructions, please sign at the bottom indicating that you have read the instructions and understand what to do. Then, place this signed page back into the envelope, seal it, place an X over the seal, and wait for the examiner to return.

You are about to take some cognitive tests that examine mental abilities such as attention, concentration, problem-solving, and your ability to think quickly. Your task is to perform to the best of your ability, answering all questions as honestly as possible while putting forth your best effort. **If it is clear that you are putting forth your best effort and trying hard on all the tests, you will receive two additional credits for a total of 6 credits.**
Please sign your name at the bottom of this page, place it back into the envelope, seal the envelope, and place an X over the seal before the researcher returns. **Remember, do not let the researcher know what you have been told to do. Your performance on the tests should be the only way of figuring it out.**

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

_____

Signature and Date

Appendix F: Role-Play Termination Instructions

If you received instructions to pretend like you sustained a concussion, at this point, please **stop** following those instructions. From this point forward, **all** participants please provide your honest and actual responses to all questions.

Appendix G: Post-test Questionnaire 1 (PTQ1)

1. Please summarize the instructions you read at the beginning of this experiment.

   _____

   _____

   _____

2. Please indicate how hard you tried to follow the instructions you were given at the beginning of the experiment by circling the one number that best describes your effort.

   1 2 3 4 5
   Didn't try at all Tried moderately hard Tried very hard

3. Please indicate how successful you think you were in producing the results asked of you in the instructions by circling the one number that best describes your success.

   1 2 3 4 5
   Not at all successful Somewhat successful Very successful

4. Do you think you were successful in keeping the examiner from discovering what your instructions told you to do?

   Yes_____ No_____

5. Please indicate how familiar you are with the effects that are often associated with a concussion by circling the number that best describes your familiarity.

   1 2 3 4 5
   Not at all familiar Somewhat familiar Very familiar

6. What do you think the test with 50 different pictures in a booklet was designed to measure? Please write only one purpose for the test.

   _____

7. What do you think the test where you read different words aloud was designed to measure? Please write only one purpose for the test.

   _____

8. What do you think the test with different numbers in circles (connected in a dot-to-dot fashion) was designed to measure? Please write only one purpose for the test.

_____

9. What do you think the test with different numbers and letters in circles (connected in a dot-to-dot fashion) was designed to measure? Please write only one purpose for the test.

_____

10. What do you think the test that asked you to quickly draw a line through all shapes of a certain type was designed to measure? Please write only one purpose for the test.

_____

11. What do you think the test where you looked at pictures on a computer was designed to measure? Please write only one purpose for the test.

_____

Thank you.

Questionnaire adapted from Huskey (2002).

Appendix H: Post-Test Questionnaire 2 (PTQ2)

It is possible that some of the tests you took today were designed to detect if someone is faking brain damage, while others are tests typically administered to assess cognitive abilities such as attention, concentration, memory, and speed of information processing.

Please put a check by any test that you took today that seemed as if it were designed to detect whether someone is faking brain damage. **IF you mark a test,** please indicate how certain you are that the test was designed to detect faked brain damage by circling the number that best describes your certainty.

_____ 50 pictures in a booklet

1 2 3 4 5
Not at all certain Somewhat certain Very certain

_____ Reading words aloud

1 2 3 4 5
Not at all certain Somewhat certain Very certain

_____ Connecting numbers in circles in a dot-to-dot fashion

1 2 3 4 5
Not at all certain Somewhat certain Very certain

_____ Connecting numbers and letters in circles in a dot-to-dot fashion

1 2 3 4 5
Not at all certain Somewhat certain Very certain

_____ Drawing a line through shapes on a page

1 2 3 4 5
Not at all certain Somewhat certain Very certain

_____ 45 pictures on a computer

1 2 3 4 5
Not at all certain Somewhat certain Very certain

Thank you.

Questionnaire adapted from Husky (2002).

Appendix I: Debriefing Statement

Thank you for participating in this study. Throughout the course of this experiment, you may have had questions regarding the nature or purpose of this study. If you still have these questions, the experimenter will be glad to answer them for you at this time.

The purpose of this study was to investigate simulated cognitive test performance following a brain injury. Data collected during the study will facilitate the development of a new neuropsychological test. Your answers to these questions, as well as your performance on the testing measures, will be kept completely confidential.

If you experienced a significant amount of discomfort during the course of the experiment, please address your concerns to the experimenter at the present time. If you feel uncomfortable doing so, you may contact the faculty supervisor of the project, Dr. Stuart Hall, at 243-4521. If you experience significant discomfort and would like to explore counseling or mental health services, students can be seen at the Clinical Psychology Center, at 243-2367 or at Counseling and Psychological Services through the Curry Health Center, at 243-4711.

IMPORTANT:

We request that you not discuss the details of this experiment with anyone who may be a future participant in the study. Thank you for your cooperation.

Appendix J: PVMT Technical Specifications

Hardware requirements for a basic desktop computer:

- Minimum 2GHz Central Processing Unit
- Minimum 4GB RAM
- Minimum 500GB hard disk space
- Monitor
- Keyboard
- Mouse

Software requirements for a basic desktop computer:

- Microsoft Office 2016 containing Microsoft PowerPoint

Images were taken using a variety of cameras including cell phones.

Image sizes and position on trial 1 (2-choice) (All image sizes and positions measured in inches from top left corner).

- Target Image Size: 11.15cm x 14.86cm
- Target Image Position: 24.13cm by 4.70cm
- Recognition Images Size: 10.39cm x 13.84cm
- Recognition Image A Position: 1.93cm by 4.09cm
- Recognition Image B Position: 18.26cm by 4.09cm

Image sizes and placement on trial 2 (3-choice)

- Target Image: 11.15cm x 14.86cm
- Target Position: 9.50cm x 4.70cm
- Recognition Image Size: 7.19cm x 9.58cm
- Recognition Image A Position: 2.57cm x 6.45cm
- Recognition Image B Position: 12.80cm x 6.45cm
- Recognition Image C Position: 23.19cm x 6.45cm

Image sizes and placement on trial 3 (4-choice)

- Target Image: 11.15cm x 14.86cm
- Target Position: 9.50cm x 4.70cm
- Recognition Image Size: 7.95cm x 11.30cm
- Recognition Image A Position: 4.67cm x 1.04cm
- Recognition Image B Position: 4.67cm x 10.87cm
- Recognition Image C Position: 18.14cm x 1.04cm
- Recognition Image D Position: 18.14cm x 10.87cm

Appendix K: Code Book for PTQ2 questions 6-12

Code Name	Code Definition	Code Includes/Key Terms
1. Memory	Use this code for any task that has to do with memory.	Recall, recognition, memory, remembering, reference to memory structures in the brain, etc.
2. Speed	Use this code for any task that has to do with speed.	Fastness, quickness, speed, thinking fast, eye speed, quick thinking, processing, reaction time, etc.
3. Attention/Concentration	Use this code for any task that has to do with attention or concentration.	Attention, concentration, focus, multi-tasking, doing many things at once, divided attention, keeping track of multiple things at one time, etc.
4. Speech/Language	Use this code for any task that has to do with speaking, talking, and reading.	Saying things, talking, repeating things aloud, saying words, writing down words, reading words, pronunciation, etc.
5. Motor	Use this code for any task that has to do with motion motor functioning.	Motion, moving, dexterity, agility, fine motor skills, gross motor skills, motor cortex, hand/eye coordination, etc.
6. Planning, Organizing, Problem-solving (POP)	Use this code for any task that has to do with executive functioning like planning, organizing, and problem solving.	Higher order thinking, reasoning, problem-solving, figuring out a problem, organizing information, sorting things out, knowing what to do, etc.
7. Comprehension	Use this code for any task that has to do with comprehension.	Understanding, knowing what things mean, comprehending, coherency, etc.
8. Other	Use this code for any task that does not fall into any category.	Relations, matching, simple tasks, schooling test, basic function, patience, effect, mental ability etc…

Examples:

- "How well you could recall things" should be coded as memory.
- "Cognition speed" should be coded as speed.
- "Verbal agility" should be coded as speech/language (and not motor).
- "Motor function" should be coded as motor.
- "Critical thinking," and "problem solving" can be coded as POP
- "Comprehension," and "understand things" can be coded as Comprehension
- "Mirroring," and "Relations" should be coded as other.
- Phrases like "mental ability" or "cognitive ability" can be coded as other since they are so vague.
- If a phrase has more than one code, choose the first code that is written.

Appendix L: Inter-rater Reliability

TOMM
Agreements 61
Disagreements 1

Word reading
Agreements 45
Disagreements 17

Trails A
Agreements 45
Disagreements 17

Trails B
Agreements 48
Disagreements 14

AST
Agreements 45
Disagreements 17

KBNA Cancellation
Agreements 49
Disagreements 13

PVMT
Agreements 62
Disagreements 0

Total

Agreements 355 (61+45+45+48+45+49+62)

Disagreements 79 (1+17+17+14+17+13+0)

Rule: 80% agreement on 95% of codes

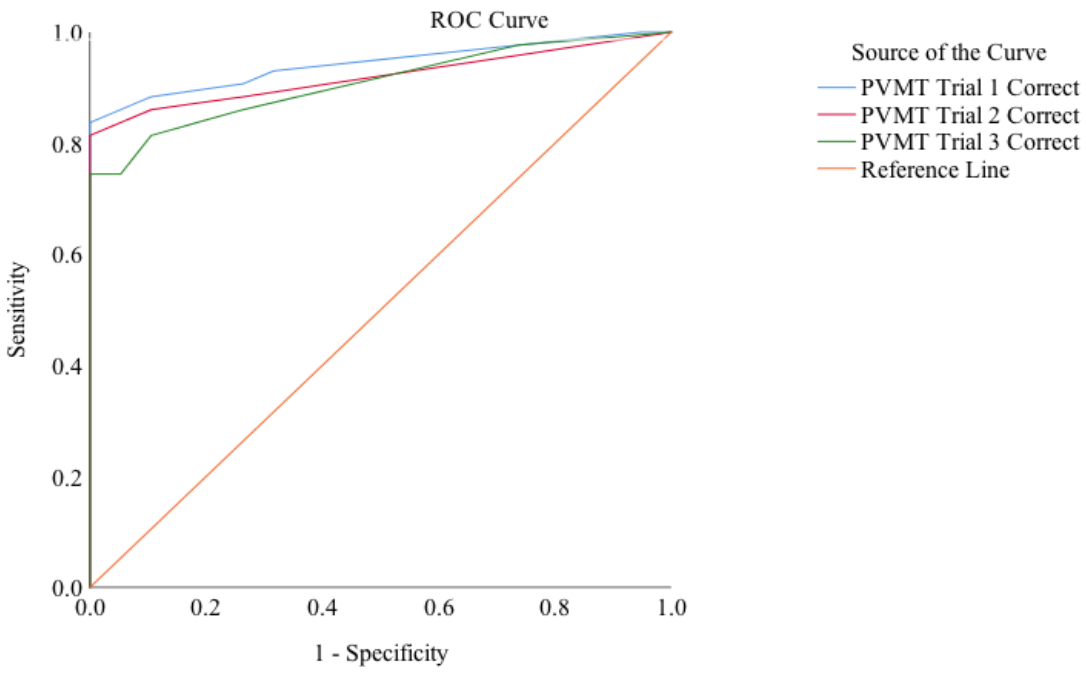
5% of 434 responses = 21.7 responses
355/355+ (79-21.7)
355/355+ 57.3
355/412.3
0.861
86.1%
86.1% agreement on 95% of codes

-OR-

82.0% agreement on 100% of codes

$$\frac{\mathbf{355}}{\mathbf{355+79}} = .\mathbf{81797} = .\mathbf{818} = \mathbf{81.8}\%$$

Appendix M: PVMT ROC Curves, AUC table, and sensitivity/specificity coordinates



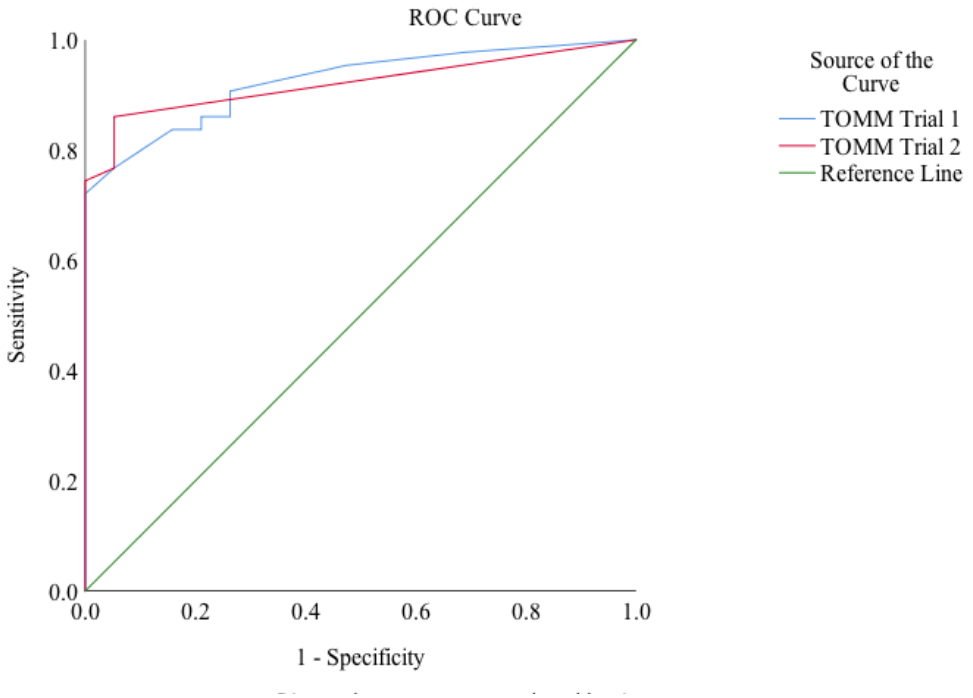
AUC for each PVMT ROC Curve

				Asymptotic 95% Confidence Interval	
Test Result Variable(s)	**Area**	Std. Error[a]	Asymptotic Sig.[b]	Lower Bound	Upper Bound
PVMT Trial 1 Correct	**.942**	.029	.000	.886	.999
PVMT Trial 2 Correct	**.920**	.035	.000	.851	.988
PVMT Trial 3 Correct	**.908**	.036	.000	.836	.979

Appendix M (cont'd): PVMT sensitivity/1-specificity coordinates

PVMT Trial 1			PVMT Trial 2			PVMT Trial 3		
Score	**Sens.**	**1-Spec.**	**Score**	**Sens.**	**1-Spec.**	**Score**	**Sens.**	**1-Spec.**
5.00	.000	.000	2.00	.023	.000	2.00	.023	.000
8.50	.023	.000	5.50	.047	.000	5.00	.047	.000
11.50	.047	.000	8.00	.070	.000	8.00	.093	.000
14.50	.070	.000	11.00	.093	.000	11.50	.140	.000
17.50	.116	.000	13.50	.116	.000	13.50	.163	.000
19.00	.140	.000	15.00	.163	.000	14.50	.186	.000
20.50	.185	.000	16.50	.209	.000	16.00	.209	.000
21.50	.209	.000	18.50	.233	.000	17.50	.256	.000
22.50	.233	.000	20.50	.256	.000	19.00	.279	.000
23.50	.302	.000	21.50	.279	.000	20.50	.302	.000
24.50	.326	.000	22.50	.326	.000	21.50	.349	.000
25.50	.349	.000	23.50	.349	.000	22.50	.442	.000
27.50	.395	.000	24.50	.395	.000	23.50	.465	.000
30.50	.465	.000	26.00	.419	.000	25.00	.488	.000
32.50	.558	.000	28.50	.465	.000	26.50	.535	.000
33.50	.581	.000	30.50	.558	.000	29.00	.558	.000
34.50	.605	.000	32.00	.581	.000	31.50	.581	.000
35.50	.628	.000	33.50	.605	.000	32.50	.605	.000
36.50	.651	.000	34.50	.628	.000	34.00	.628	.000
38.00	.721	.000	36.00	.674	.000	36.50	.674	.000
39.50	.814	.000	38.50	.698	.000	38.50	.698	.000
40.50	.837	.000	40.50	.721	.000	40.00	.744	.000
41.50	.884	.105	41.50	.767	.000	41.50	.744	.053
42.50	.907	.263	42.50	.814	.000	42.50	.814	.105
43.50	.930	.316	43.50	.860	.105	43.50	.860	.263
44.50	1.000	.947	44.50	.884	.263	44.50	.977	.737
46.00	1.000	1.000	46.00	1.000	1.000	46.00	1.000	1.000

Appendix N: TOMM ROC Curves, AUC table, and sensitivity/1-specificity coordinates



AUC for each TOMM ROC Curve

Test Result Variable(s)	**Area**	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
TOMM Trial 1	**.924**	.032	.000	.861	.988
TOMM Trial 2	**.921**	.035	.000	.853	.989

Appendix N (cont'd): TOMM sensitivity/1-specificity coordinates

TOMM Trial 1			TOMM Trial 2		
Score	**Sens.**	**1-Spec.**	**Score**	**Sens.**	**1-Spec.**
9.00	.000	.000	.00	.000	.000
14.00	.023	.000	6.50	.023	.000
19.00	.047	.000	12.50	.047	.000
20.50	.093	.000	13.50	.070	.000
21.50	.140	.000	16.00	.093	.000
22.50	.186	.000	19.50	.116	.000
23.50	.233	.000	22.00	.140	.000
24.50	.279	.000	23.50	.209	.000
25.50	.326	.000	24.50	.233	.000
26.50	.372	.000	25.50	.279	.000
27.50	.395	.000	26.50	.326	.000
28.50	.419	.000	28.00	.395	.000
29.50	.442	.000	30.50	.442	.000
30.50	.465	.000	33.00	.465	.000
32.50	.512	.000	34.50	.535	.000
34.50	.558	.000	35.50	.558	.000
36.00	.581	.000	36.50	.581	.000
37.50	.605	.000	38.50	.605	.000
39.00	.674	.000	41.00	.651	.000
40.50	.721	.000	43.50	.674	.000
41.50	.767	.053	45.50	.744	.000
42.50	.837	.158	46.50	.767	.053
43.50	.837	.211	47.50	.814	.053
44.50	.860	.211	49.00	.860	.053
45.50	.860	.263	51.00	1.000	1.000
47.00	.907	.263			
48.50	.953	.474			
49.50	.977	.684			
51.00	1.000	1.000			