DIAGNOSTIC AND CONSTRUCT VALIDATION OF SYMPTOM AND PERFORMANCE VALIDITY TESTS OF MALINGERING IN A CIVIL LITIGATION CONTEXT

A Thesis

Presented to

The Faculty of the Department

of Psychology

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

By

Katherine A. Fox, M.A.

August, 2018

DIAGNOSTIC AND CONSTRUCT VALIDATION OF SYMPTOM AND

PERFORMANCE VALIDITY TESTS OF MALINGERING

IN A CIVIL LITIGATION CONTEXT

Katherine A. Fox, M.A.

APPROVED:

John P. Vincent, Ph.D.
Committee Chair
Department of Psychology
University of Houston

Gerald Harris, Ph.D.
Department of Psychology
University of Houston

Christopher Barr, Ph.D.
Texas Institute for Measurement, Evaluation, and Statistics
University of Houston

Antonio D. Tillis, Ph.D.
Dean, College of Liberal Arts and Social Sciences
Department of Hispanic Studies

DIAGNOSTIC AND CONSTRUCT VALIDATION OF SYMPTOM AND

PERFORMANCE VALIDITY TESTS OF MALINGERING

IN A CIVIL LITIGATION CONTEXT

An Abstract of a Thesis

Presented to

The Faculty of the Department

of Psychology

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

By

Katherine A. Fox, M.A.

August, 2018

ABSTRACT

The extent to which persons may malinger psychiatric symptoms is a legitimate concern in civil litigation. The consequences inherent in personal injury cases involving psychological distress necessitate an understanding of how malingering presents in medico-legal contexts and the validity and usefulness of available methods to detect malingering. The present study evaluated the construct and diagnostic validity of symptom-based (SVT) and performance-based (PVT) measures of malingering in a simulated personal injury paradigm. We evaluated the interrelationships between malingering measures and whether these measures were able to discriminate between "honest responders" and "malingerers." Using a simulation design, 411 undergraduates were randomly assigned into four experimental conditions, which outlined the experience of a motor vehicle accident and subsequent psychological and cognitive symptoms. Conditions varied on the degree of suggestion to malinger symptoms as related to a personal injury case. Under this paradigm, participants completed measures of malingered symptomatology, including the TOMM, M-FAST, SIMS, and TSI-2 ATR. As predicted, we found weaker correlations between PVT and SVTs, but moderate significant correlations were found across symptom validity measures. These findings support conceptualization of malingering as a non-unitary construct. Results from ROC analysis suggest that only the TSI-2 ATR was useful in discriminating between simulation groups. Contrary to expectations, prominent measures of malingering (TOMM, M-FAST, and SIMS) did not discriminate between groups. Results may mean that these tests may operate differently than intended within a civil litigation context and depending on the type of psychopathology feigned.

TABLE OF CONTENTS

Diagnostic and Construct Validation of Symptom and Performance Validity Tests of Malingering in a Civil Litigation Context

Personal Injury Litigation

Personal injury lawsuits are a form of civil litigation which allow individuals to claim monetary restitution on the basis of a breach of care or civil wrong, also known as a tort, as brought about by the actions or inactions of others. Examples of torts include premises liability, medical and professional malpractice, product liability, assault, defamation, invasion of privacy, and personal injury as the result of an accident that otherwise would have been prevented, if not for the tortfeasor's conduct (Melton, Petrila, Poythress, & Slobogin, 2007; Cohen, 2004). Personal injury claims can result from motor vehicle accidents, in which the defendant's negligent or reckless driving behavior lead to either physical or psychological harm. While the criteria for types of torts differ, related claims require the breach of an established duty on the part of the defendant, which proximately causes "damages," or a compensable injury. Through the process of personal injury litigation, individuals injured in such events (the plaintiffs) are able to sue for monetary damages and can be awarded financial compensation from those held responsible (the defendants; Foote & Lareau, 2013). Damages sought in personal injury cases can include pain and suffering; past and future economic losses; or medical expenses, including psychological services, incurred in attempts to alleviate symptoms of mental or physical injury. According to the Bureau of Justice Statistics, in a 2005 survey of civil court cases in the United States, 57% of all tort claims (N = 16,397) were automobile accident cases (Langton & Cohen, 2008). Taking into consideration that only approximately 4% of these civil cases go to trial, as the vast majority are settled out of court, it is estimated that over 400,000 personal injury claims were made in 2005.

In order to make a successful personal injury claim, proof must be provided that the defendant behaved in a negligent or reckless manner (either by action or inaction) and that the resulting injuries would not have occurred if it were not for the defendant's behavior (Young, Kane, & Nicholson, 2007). This requirement, known as causation, holds that the defendant's conduct must be reasonably associated with the resulting injury; however, it need not be the sole contributory factor. As noted by Ackerman and Kane (1998), actions carried out by the defendant need only play a role in initiating, perpetuating, or aggravating injuries sustained, to qualify as a proximate cause and sufficient grounds for a personal injury claim. As such, there are a variety of conditions which may lead to liability in a personal injury context.

Psychological or Mental Injury

In addition to physical injuries sustained as the result of a tortious act, individuals may experience psychological injuries, which can range from mild, transient experiences such as embarrassment to severe, chronic conditions such as depression and trauma-related mental disorders (Vallano, 2013). For example, individuals involved in serious motor vehicle accidents may experience a host of psychological and functional sequelae as a result of the accident. These sequelae can include anxiety, depression, post-traumatic symptoms, and the inability to maintain occupational and social functioning (e.g., being unable drive to work or school, social isolation as a result of avoidance of driving; Foote & Lareau, 2013). In the United States, a majority of jurisdictions permit damages for emotional distress or mental injury, in addition to and in the absence of physical harm (Foote & Lareau, 2013). It is estimated that 50% of civil injury awards involve the experience of psychological pain and suffering (Vallano, 2013), reflecting the substantial impact that personal injury claims have on the civil court system.

Claims of cognitive injury or emotional harm are often disputed by defendants in personal injury cases. As such, the law requires “objective indicia of mental injury” (Melton, Petrila, Poythress, & Slobogin, 2007), often in the form of an official diagnosis or professional opinion that mental injury has occurred, as a result of the precipitating event. This necessity places psychologists at the forefront of many mental injury cases, as they lend significant credibility to the claim. Psychologists are thus tasked with evaluating the presence and veracity of psychological injuries, as a direct result of the defendant’s actions, and determining the degree of impact that such injuries have had on claimants (Foote & Lareau, 2013). The issue of causation is noted to be particularly complex in personal injury cases involving psychological injuries, as pure and direct etiologies for mental disorders are often elusive or non-existent. Rather, “psychological causation” is considered to be multifactorial, interactive, and non-linear, which is discordant with the simplicity and linearity of legal tests (Schultz, 2003), and thus places a particular burden on psychologists to comprehensively assess psychological injury and the context in which the injury was sustained. This assessment often includes consideration of pre-existing factors (e.g., childhood trauma history), in addition to peri- and post-traumatic factors (Melton, Petrila, Poythress, & Slobogin, 2007).

Since it is possible in the U.S. Courts to receive monetary compensation for psychological injuries, the extent to which persons seeking compensation may falsify psychiatric symptoms is “a legitimate concern” in civil litigation (Peace & Masliuk, 2011).

The opportunity for financial gain through civil litigation provides substantial incentive for individuals to malinger, or fake, mental injury. An injury that is often claimed in these cases is posttraumatic stress disorder, or PTSD (Melton, Petrila, Poythress, & Slobogin, 2007). Indeed, it has been suggested that the diagnosis of PTSD is “made-to-order” for plaintiffs in personal injury cases, due to the requirement of an etiological stressor, or casual event, in the diagnostic criteria. In personal injury cases, the stressor which is claimed as the precipitant of the PTSD symptoms can involve a variety of events, including motor vehicle accidents, industrial accidents, and both physical and sexual assault (Taylor, Frueh, & Asmundson, 2007). The utilization of PTSD as a basis of personal injury claims has become so commonplace as to support an industry of attorneys who solely litigate cases involving PTSD (Stone, 1993).

PTSD: Diagnostic Criteria

Hall and Hall (2007) state directly that PTSD is “particularly vulnerable” to falsification due to its reliance on subjective symptoms and histories provided by self-report. According to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5;* American Psychiatric Association, 2013), there are five overarching categories that comprise the criteria for a diagnosis of PTSD. The first category, and the most critical, is the exposure to severe stress or trauma (Criterion A). To meet this criterion, individuals claiming the experience of PTSD must have been exposed to either actual or threatened death, serious bodily injury, or sexual violence. Notably, this exposure need not be first-hand, but may have occurred as the result of witnessing such events or learning of such events occurring to a family member or close friend (American Psychiatric Association, 2013).

After exposure to a traumatic event, individuals who meet criteria for PTSD experience a variety of symptoms which lead to either clinically significant distress or functional impairment. The experience of significant cognitive, affective, and physiological symptoms of traumatic stress must persist for at least one month. Post-traumatic symptoms, distinct from the experience of a traumatic event, fall into four primary groups, including: intrusion symptoms (Criterion B), avoidance symptoms (Criterion C), negative cognitions and mood (Criterion D), and hyperarousal (Criterion E; American Psychiatric Association, 2013). In order to meet full diagnostic criteria, individuals must experience at least one intrusion symptom, which includes experiences such as flashbacks or distressing dreams; at least one avoidance symptom, which includes behaviors such as avoiding places, people, or situations associated with the trauma; two cognitive or mood symptoms, including social detachment, depression, or extreme guilt; and two symptoms classified as hyperarousal, which includes hypervigilance, recklessness, or exaggerated startle response. Individuals may also experience persistent or recurrent dissociative symptoms, including depersonalization (detachment from one’s body or sense of self) or derealization (detachment from the external world or reality).

Prevalence of Trauma

The experience of events that can be categorized as traumatic is not uncommon in the general population. According to Norris (1992), in a sample of 250 persons drawn from Southern cities in the United States, approximately 69% of individuals experienced at least one type of traumatic event in their lifetime and 21% experienced a qualifying event during the previous year. Traumatic events in this study included sexual assault, tragic death, natural disasters, robberies, and injury-causing motor vehicle accidents. Of these events, tragic deaths, robberies, and automobile accidents were the most prevalent, with the latter having occurred to approximately one-fourth of sampled individuals across a lifetime. However, not all individuals who experience a Criterion A event will go on to develop PTSD. According to Breslau (2001), epidemiological studies suggest that only 15-24% of individuals who experience a traumatic event will develop the disorder. The percentage of individuals who go on to develop PTSD also varies as a function of the event. As noted by Resnick and colleagues (2008), it has been reported that up to 80% of individuals who experience rape will go on to develop PTSD symtomatology, but only 15-30% of those involved in motor vehicle accidents. Furthermore, a smaller fraction of those who experience symptoms will meet full diagnostic criteria for PTSD and will seek external support for the management of symptoms, including both psychological counseling and legal restitution. According to the DSM-5, the twelve-month prevalence for the development of new PTSD cases is approximately 3.5% in the United States (American Psychiatric Association, 2013).

As noted by Resnick and colleagues (2008), the identification of malingered PTSD profiles has become one of the most challenging tasks for clinicians, independent of legal context. Unlike most DSM diagnostic criteria, the symptoms that comprise Criteria B, C, D, and E of PTSD do not require external corroboration and are largely unobservable. In addition, information about PTSD symptoms is widely available to the public via the internet and print media (Purtle, Lynn, & Malik, 2016), and portrayals of persons suffering from PTSD have featured in major motion pictures and television series. This prominence aids the lay-public in understanding more fully the impact of trauma, but also increases the ability for individuals undergoing an assessment to know what to expect and how to respond to assessment questions regarding PTSD. Standard assessment measures of affective and cognitive PTSD symptoms provide little information regarding the veracity of these symptoms, unless major discrepancies between reports were found. Several studies have demonstrated that “people, regardless of their knowledge of psychopathology, are capable of feigning psychological injury” including both PTSD and depressive disorders (Arce, Farina, & Buela, 2008). In their work investigating whether naïve subjects could produce believable symptom profiles of PTSD, Lees-Haley and Dunn’s findings (1994) supported this conclusion, showing that 98.9% of subjects could successfully meet the requirements for Criterion B on self-report questionnaires, and 95.7% could successfully meet Criterion E without coaching or practicing. Thus, it is evident that the symptoms PTSD can easily be feigned regardless of the veracity of their existence, and regardless of depth of psychological knowledge and direct coaching.

Complicating the clinical picture, PTSD is a heterogeneous disorder in which true sufferers may endorse very different symptoms (Zoellner, Pruitt, Farach, & Jun, 2014). This heterogeneity increases the difficulty of discerning true versus feigned presentations, as there is no definitive model of PTSD symptomatology. Further confounding this process is the extent to which PTSD can coexist alongside other disorders, commonly known as comorbidity. The presence of fear, distress, and dysphoric symptoms, inherent in a PTSD diagnosis, reflects the extent to which PTSD can overlap with a variety of other conditions, including depressive, dissociative, anxiety, and even personality disorders (Zoellner et al., 2014). In total, the lack of distinct boundaries which demarcate PTSD symptoms from other conditions and the degree to which true symptom presentations of PTSD can vary makes discerning true PTSD from feigned presentations very difficult.

Malingering and Response Style

There are several distinct terms that are utilized when discussing response styles in forensic and general clinical practice. Response styles, such as malingering and feigning, refer to the overarching ways in which persons undergoing psychological evaluation may approach the assessment situation. The most egregious response style is malingering, which is defined as the "deliberate fabrication or gross exaggeration of psychological or physical symptoms for the fulfillment of an external goal" (American Psychiatric Association, 2000). The concept of feigning is often confused with that of malingering because it involves symptom exaggeration and fabrication without an assumption about the goal or intended outcome of the fabrication (Rogers & Bender, 2013). Strictly defined, an individual who is fabricating or exaggerating symptoms would be considered to be feigning up until the point in the assessment in which the purpose of that fabrication is made evident to the evaluator. At such a point, the individuals may be identified as malingering. It is important to highlight, however, that such a revelation of motive may never occur (Berry & Nelson, 2010). Some authors also note that psychological assessments cannot provide information about intentions or goals. Therefore, all assessment measures which address response style are considered to be measures of feigning and not malingering (Rogers & Bender, 2013).

Phillip Resnick (1997) identified three main subtypes of malingering which can be applied to PTSD. Pure malingering involves a complete fabrication or falsification of symptom presentation. With PTSD, this can include the fabrication of traumatic experiences, in addition to falsified experience of psychological sequelae. Partial malingering, alternatively, involves the exaggeration of existing symptoms or the endorsement of symptoms which are currently in remission. Partial malingering can also include the fabrication of a traumatic event or the over-dramatization of an event to traumatic levels (Taylor et al., 2007). As noted by Peace and Masliuk (2011), PTSD is particularly vulnerable to malingering because of the ability to fabricate experiences in addition to symptoms. For example, due to the often private experience of events such as domestic violence and sexual assault, it is possible for individuals to falsify the occurrence or severity of these events in order to make a personal injury claim. Lastly, there is the concept of false imputation which involves the intentional misattribution of symptoms to a traumatic experience. False imputation can also involve the legitimate experience of two traumatic events, where the symptoms associated with one event are purposely misattributed to the other in order to achieve an external goal (Resnick, 1997). Along with direct compensation afforded through damages in the civil court process, there are also many other types of external motivators which may lead plaintiffs to fabricate symptoms, such as seeking revenge, validation, or attention (Peace & Masliuk, 2011).

Deception Strategies

Evidence suggests that persons attempting to feign psychiatric distress often resort to two primary strategies, either describing their symptom severity as extremely high or endorsing as many symptoms as possible (known as indiscriminate symptom endorsement; Arce et al., 2008; Peace & Masliuk, 2011). In the context of PTSD and civil litigation, further evidence has been found to support the theory that feigners will either try to make claims of psychological distress as believable as possible (in order to guarantee compensation) or to exaggerate symptoms excessively in order to receive as much compensation as possible. According to Williams and colleagues (1999), individuals seeking monetary compensation through legal avenues for psychological and neuropsychological injuries tend to report a greater number of symptoms, and indicate that such symptoms have persisted for longer than individuals not involved in litigation.

Prevalence of Malingering

Determining the prevalence of malingering in clinical and forensic populations is quite difficult. Naturally, the hope of the individual is that his or her feigning will not be detected and it can be presumed that great efforts are made to ensure that detection does not occur. Despite these factors, many authors have attempted to estimate the occurrence of malingering in various populations. Estimates regarding feigning of psychiatric symptoms range from 1% to over 50% depending on the context of the case and the reason for symptom presentation (i.e., for insurance purposes, clinical samples, or civil litigation; Hall & Hall, 2007; Peace & Masliuk, 2011). According to estimates by forensic practitioners, malingering likely occurs in 15-17% of forensic cases (Rogers & Bender, Evaluation of Malingering and Related Response Styles, 2013). Some estimates suggest that malingering occurs in up to 40% of civil litigation cases involving neuropsychological assessment (Larrabee, 2003; Young, Kane, & Nicholson, 2007), while other studies have found that 20-30% of results from psychometric testing on personal injury plaintiffs suggest that malingering had taken place (Taylor et al., 2007). Further obscuring the prevalence of malingering is the extent to which “malingering” (or feigning) is defined and operationalized across researchers and the selection and efficacy of detection methods used to determine presence of a falsified profile.

Several authors note, however, that these estimates may be obscured by the extent to which persons may vary in their malingering throughout the process of assessment and across testing periods. Often erroneously considered a “monolithic” and stable construct, Berry and Nelson (2010) point out that persons who malinger may be inconsistent in their false responding and that there are at least three separate targets of malingered symptoms: psychiatric, physical and somatic, and cognitive/neuropsychological. It cannot be assumed that known feigners will always falsify responses in assessment, or that individuals who feign one type of symptom will also feign others (Rogers, 2008). Similarly, individuals undergoing assessment are subjected to other factors which may influence responding and operate independently from intention to feign. Responding is additionally influenced by fatigue, cognitive capabilities, and distractions, among other factors, which may result in suboptimal performances or invalidated symptom profiles that are not reflective of a willful fabrication of symptoms. Lastly, it is valuable to note that these estimates may not take into account the fact that the feigning does not rule out the possibility of true symptoms existing within the individual (Rogers & Bender, 2013). As noted by researchers Guy, Kwartner, and Miller (2006), the feigned disorder is often a function of context as well. For example, individuals may be more likely to feign psychosis in criminal litigation settings, whereas feigned PTSD and other affective conditions may be more associated with civil litigation proceedings.

Assessment and Detection of Malingering

As outlined above, through the civil court process and by means of personal injury litigation, there exists a substantial incentive for individuals to feign psychiatric symptoms as a result of a traumatic event. Furthermore, the diagnosis of PTSD lends itself well to claims of this kind, due to its requirement of an etiological stressor and lack of objectively-determined symptom presentations. Because of this incentive to malinger, it is the role of mental health professionals to verify the existence and severity of PTSD symptoms, and to lend credibility to claims of psychological pain and suffering. Several psychologists have attempted to develop detection strategies by means of traditional assessment methods

(Rogers & Bender, 2013; Berry & Nelson, 2010; Thomas & Fremouw, 2009). Rogers and Bender (2013) have investigated detection strategies extensively and have identified several approaches which can be categorized in two main ways: unlikely presentations or amplified/improbable presentations (Rogers & Bender, 2013). In the series of unlikely detection strategies, clinicians are encouraged to focus on those symptom endorsements which are considered rare, improbable, or erroneous, and to note unusual symptom combinations and spurious patterns. Thus, in this category, the presence of certain symptoms is the primary indicator. In the series of improbable or amplified detection strategies, the focus of clinical attention is that of the magnitude, severity, or obviousness of symptoms endorsed. As such, this category focuses on the magnitude of symptom endorsements as the primary indicator of feigning. The theoretical basis of these strategies is the assumption that those who malinger are unlikely to have sufficient knowledge of the disorder and general psychopathology to an extent that they can accurately portray its symptoms. As a result, malingerers are deemed likely to endorse unusual or, at times, impossible symptom combinations and will often exaggerate the severity of symptoms over and above the severity which is found in true clinical populations. Extensive programs of research into the assessment of malingering have provided tools for detecting these falsified symptom presentations (Rogers & Bender, 2013).

Symptom Validity Tests

Symptom validity tests (SVTs) are measures which aim to detect the exaggeration or fabrication of psychiatric symptoms, based on self-report endorsements of experience. SVTs utilize a variety of detection methods, which often capitalize on the relative infrequency, atypical combination, or usual severity of reports of psychological symptoms. In attempts to develop valid and reliable methods of detecting malingering or feigned mental illness via traditional assessment, two main psychometric approaches have emerged: large, multi-scale inventories and brief, domain-specific measures (Guy, Kwartner, & Miller, 2006; Rogers, 2013).

The first of these approaches includes the utilization of validity and response style scales embedded into lengthy, self-report measures of personality and psychopathology. The most commonly used and recognized methods in this class of approaches are the Minnesota Multiphasic Personality Inventory – 2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kreammer, 1989) and the Personality Assessment Inventory (PAI; Morey, 2007, Boccacini & Brodsky, 1999). In these assessments, examinees respond to a series of questions (true or false in the MMPI; Likert scale on the PAI) which attempt to garner information about symptom presentation. Endorsement and denial patterns are mapped onto clinical scales, which reflect a broad array of psychiatric experiences. Alongside these clinical scales, validity scales are included which aim to assess inconsistent responding and factors such as defensive responding and the exaggeration of symptoms. Extensive studies have been conducted regarding the validity of these measures for the purposes of malingering detection, including feigned PTSD (Resnick, West, & Payne, 2008).

A primary advantage of large-scale assessment measures, such as the MMPI or PAI, is the ability to compare symptom profiles with a normed population and therefore to determine sufficient cut-off scores based on clinically significant elevations. Unfortunately, with PTSD, there has been much controversy regarding the appropriate cut-off scores to meet symptom criteria. In particular, it has been found that the results of the majority of true PTSD sufferers included elevations on most scales in comparison with sufferers of other true disorders. For example, one study found that in a clinical sample of victims of adult sexual abuse, 20% of respondents on the MMPI-2 Infrequency (*F)* scale had a T-score greater than 100 (5 standard deviations above the mean; Klotz Flitter, Elhai, & Gold, 2003). Thus, researchers caution the use of strict cut-off scores, as it likely that even patients with genuine PTSD symptoms may produce elevated or exaggerated profiles (Hall & Hall, 2007). As discussed by Klotz-Flitter and colleagues (2003), significant elevations on malingering scales (e.g., the F scale on the MMPI-2) may be reflective of severe, genuine pathology or distress for trauma victims. Elevations may result from conscious or unconscious attempts to "cry for help," or may result from dissociative experiences, in addition to typical PTSD symptomatology, which lead to atypical or "chaotic", but genuine, symptom profiles.

Disadvantages of these methods include the degree to which they are time-consuming, dependent upon minimum reading proficiencies, costly, and, most importantly, susceptible to coaching (Guy, Kwartner, & Miller, 2006). Since these measures are considered to be direct, meaning that they seek endorsements of specific symptoms known to make up a variety of disorders, self-report inventories are considered to be easily faked (Thomas & Fremouw, 2009). One study investigating this susceptibility in the MMPI-2 has found that one-third of people were able to elude detection of malingering psychotic symptoms when they were coached about the validity scales. As such, clinicians in forensic settings must be cognizant to the possibility of coaching, and often turn to other methods of assessing malingering.

Bridging the gap between larger inventories and domain-specific measures of malingering, are measures such as the Trauma Symptom Inventory (TSI-2; Briere, 2011), which holds a narrower focus on the symptoms and experiences common to traumatic stress disorders, but also includes embedded validity subscales, such as the Atypical Responding scale (ATR). As with other large-scale inventories, an advantage of the TSI-2 and its predecessor, the TSI (Briere, 1995), is the ability to identify cut-scores based on comparisons with various clinical and community populations. While valuable in its relative specificity toward trauma experiences, Resnick and colleagues (2008) note that the original TSI, and more specifically the ATR subscale, demonstrated only marginal ability to differentiate between genuine PTSD patients and malingerers, leading to a substantial false positive rate (i.e., individuals incorrectly identified as malingering; Elhai, et al., 2005). Authors Elhai and colleagues (2005) observed that the TSI ATR subscale was comprised of ten items which reportedly address bizarre, unusual and psychotic experiences, which were not selected based on a infrequency criterion (e.g., identified as occurring in less than 10% of the standardization sample) and were based on the responses of asymptomatic healthy individuals. Based on their findings, clinicians were strongly cautioned against the sole use of the TSI ATR scale to determined malingered PTSD. While substantial limitations to the use of this measure were noted, Boccaccini and Brodsky (1999) found that in a survey of 80 psychologists involved in emotional injury cases, approximately 33% reported the use of the TSI in addition to other measures in their assessment batteries. A more recent survey of test usage found that in an international sample of 868 assessment cases conducted by 434 clinicians, approximately 21% involved the use of the TSI in civil tort cases, reflecting the second most commonly used measure (Neal & Grisso, 2014). Furthermore, in a more recent analysis, researchers Gray, Elhai, and Briere (2010) found that the TSI-2 ATR scale was able to correctly classify 74.2% of known-groups comparison sample into simulators and genuine PTSD sufferers (when using a cut score of 7).

In addition to large, multiscale inventories which include validity scales, there are several specific, "standalone" measures which attempt to discern malingering or symptom exaggeration more briefly and directly (Parks, 2015). Some examples of these measures include: the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992); the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001); and the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005). Domain-specific measures such as these boast strong psychometric properties and have been found to be used frequently in practice (Guy et al., 2006; Roger & Bender, 2013). Multiple studies show support for the use of these standalone SVTs in the detection of feigned amnesia, epilepsy, and psychosis (Parks, 2015).

Similar to the multiscale inventories, these measures are both direct and based on self-report. Unlike the MMPI-2 and PAI, however, many of these measures (such as the SIRS and M-FAST) are structured interviews which do not rely on minimum reading proficiencies and thus may be better suited for forensically-involved populations (Guy et al., 2006). One key advantage of the structured interview format is also the assumption that malingerers will find it more difficult to lie face-to-face (as when interviewed in person), as opposed to lying on paper (while taking a paper-and-pencil test).

Another primary difference is the extent to which these measures were designed for the detection of feigning and the extent to which they address specific disorders. While this specificity can be valuable to the clinician, it can also lead to difficulties in detecting feigning when the clinician is not aware of the disorder specific nature of a particular assessment. For example, the M-FAST has been found to address more heavily symptoms associated with, or believed to be associated with, psychotic experiences, along with "highly unusual" and "improbable symptoms" (Guy et al., 2006). Authors have thus questioned the effectiveness of the M-FAST in detecting malingering across a larger spectrum of potential disorders (Vitacco, Rogers, Gabel, & Munizza, 2007). Similar evaluations have been made regarding the SIRS, which a highly-structured measure reportedly designed to evaluate the veracity of reported psychotic-spectrum symptoms (Melton, Petrila, Poythress, & Slobogin, 2007). Studies investigating the extent to which various assessment measures are used in emotional injury cases found that 26% of examiners have reported using the SIRS, despite the noted infrequency with which symptoms of psychosis are likely to present in emotional injury contexts (Boccaccini & Brodsky, 1999; Melton et al., 2007; Guy et al., 2006). Additionally, it is noted by some authors that these measures may be more effective when attempting to determine the falsified presentation of general psychopathology, as opposed to the false presentation of a specific disorder (Guy, Kwartner, & Miller, 2006). In light of these limitations, several authors stress the need for clinicians to be cognizant of the characteristics of the measures selected and the extent to which such measures are appropriate for the given situation, presentation, and individual (Berry & Nelson, 2010; Rogers & Bender, 2013).

Performance Validity Tests

In contrast to symptom validity tests, which aim to verify symptom report based on endorsement patterns, performance validity tests (PVTs) or effort-based measures, are designed to verify symptom presentation based on actual examinee performance on various tasks of neurocognitive functioning. Similar to SVT's, PVT's can exist as standalone measures or as embedded indicators within larger test batteries. Also like SVTs, PVTs utilize a variety of detection methods, including identifying uncommon or unlikely performance presentations, when compared to a genuinely impaired normative sample, and a phenomenon known as the *floor effect.* The floor effect operates under the principle that malingerers will feign significant impairment on tasks which even truly impaired individuals would be able to pass (Rogers, 2008). The Test of Memory Malingering (TOMM; Tombaugh, 1996), is one measure that uses such a strategy. This task, which requires examinees to recall a series of fifty line drawings, appears difficult; however, in a validation sample of 145 patients with confirmed neurological impairment, reflecting four diagnostic categories (cognitive impairment, aphasia, traumatic brain injury, and dementia), all diagnostic groups averaged above the recommended cut-off score of 45 out of 50 items correctly recognized (Tombaugh, 2006). According to the author, the cut-off score of 45 on the TOMM trial 2 correctly classified (i.e., identified as non-malingering, true patients) 100% of community-dwelling participants and 95% of non-demented participants. As such, the TOMM operates under the assumption that individuals scoring fewer than 45 correct out of 50 must be exaggerating or falsifying impairment, and therefore warrant further evaluation. Notably, the TOMM has been found to be the most frequently used measure to detect poor effort or malingering among a sample of 188 neuropsychologists (Sharland & Gfeller, 2007).

Despite extensive research, no standardized protocol for detecting malingering has been developed or validated. According to Neal and Grisso (2014), the average number of measures used by psychologists conducting evaluations in civil tort cases was 4.6, with some cases using up to eighteen different measures. Researchers Boccaccini and Brodsky (1999) noted that, in their survey of 80 emotional injury evaluators who conducted a career sum of 10,500 evaluations, “no two practitioners used the exact same combination of tests.” As noted above, there are a variety individual measures of malingering and symptom validation which each hold strengths and weaknesses that influence a clinician’s decision to employ a given measure. Due the availability of measures and their various limitations (e.g., minimum reading level, difficulty of administration, cost, time-efficacy, comprehensiveness), clinicians are often encouraged to employ a combination of methods and information sources, including multi-scale inventories with imbedded validity indicators, domain-specific measures of malingering, assessments of general cognitive and neuropsychological functioning, behavioral observations and clinical interviews, and collateral data via records and third-person reports (Resnick, West, & Payne, 2008). By integrating and comparing these various sources of information, clinicians are asked to evaluate inconsistencies and inaccuracies in an individual's report of symptoms, in addition to clinically significant elevations on various malingering indictors, the presence of which in sum may be suggestive of feigning. As a result, a major research area in the field of malingering detection has been the evaluation of assessment methods and developing a greater understanding of the degree to which various measures of malingering and peripheral functions (e.g., cognitive functioning, memory functioning, psychiatric symptom influence) operate interdependently. Due to the significant legal and financial ramifications of such evaluations, it is crucial that clinicians employ empirically validated and reliable instruments to make their decisions, and that clinicians understand the construct of malingering and malingering assessment such that decisions based on these data are empirically supported.

Construct Validity and Multitrait-Multimethod Matrices

Construct validity, colloquially known as the extent to which a test measures what it is intended to measure, involves the incorporation of various evidentiary points, which in sum indicate a valid assessment of the proposed construct. As noted in the Standards for Educational and Psychological Testing (SEPT; American Educational Research Association , 2014), while prior conceptualizations purported validity to be a distinct feature of a test, more recent conceptualizations have emphasized the extent to which validity is not a "static property," but rather a dynamic, "evaluative judgement" (Goodwin & Leech, 2003; Messick, 1995). Under this interpretation, measures cannot *have* construct validity, but rather clinicians and researchers *make an argument for* construct validity based on available information. One primary method considers the extent to which a measure correlates with other tests of that kind (convergent validity) and does not correlate with other tests deemed to be unassociated (discriminant validity; Campbell & Fiske, 1959). By evaluating inter-correlations amongst measures, researchers can investigate whether the measures included in their study conform to this rule, and thereby supply evidence for construct validity.

Campbell and Fiske (1959) noted that there are additional factors which contribute to inter-correlations amongst measures, beyond measuring the same construct. Method variance is variance which is attributable to the means by which information is gathered, rather than the constructs intended to be assessed (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Method variance can include how the measure was administered (e.g., paper-and-pencil task vs. computer vs. interview format) or who is completing the measure (e.g., self-report vs. other-report). As such, evidence of discriminant and convergent validity requires the evaluation of multiple traits, assessed through multiple methods. An array of inter-correlations, known as a multitrait-multimethod matrix (MTMM; Campbell & Fiske, 1959), allows for researchers interested in validating measures to examine whether assumptions of construct validity hold, while considering the influence of method variance.

In a MTMM matrix, convergent validity is evaluated via strong inter-correlations amongst measures which are proposed to assess the same underlying construct, while utilizing the same method. Additional support for the validity of a given measure is found via inter-correlations amongst other measures of the same construct, which use different methods. Divergent validity is evaluated via weak or negative correlations amongst measures that use similar methods to assess different constructs. Broadly, construct validity is evaluated and supported through the demonstration that measures relate with other measures of the same construct in a manner that is consisted with proposed theoretical relationships.

Limited research has been conducted using MTMM relationships in support of validity for SVT malingering measures. Story (2000) used a MTMM matrix in order to evaluate the inter-relationships between the MMPI-2 and the Abbreviated SIRS on scales relevant to malingering and response style. In this study, convergent validity was demonstrated through significant correlations between SIRS and MMPI-2 malingering indices, whereas divergent validity was indicated via lower or non-significant correlations between the SIRS and non-malingering MMPI-2 indices. Results were consistent with researcher expectations, with all SIRS scales having a significant, positive correlation with MMPI-2 malingering indices (e.g., F, F-K, Fb, $Ds\text{-}r_2$, and F(p)), and holding non-significant and low correlations with MMPI-2 indices of defensiveness (L and K). In this study, the MTMM matrix was used to provide validity evidence for the use of an abbreviated version of the SIRS, a widely known SVT measure of malingering.

Several researchers have highlighted the importance of utilizing both PVTs and SVTs in the evaluation of malingered neurocognitive impairment, specifically traumatic brain injury. Some authors have extended this recommendation to include feigned posttraumatic stress disorder, which can co-occur with the experience of a brain injury. Two authors have examined the relationship between PVT and SVT failure. Demakis, Gervais, and Rohling

(2008) found that elevated psychological symptoms were not associated with PVT failure, nor were poorer performances on measures of neuropsychological functioning associated with SVT failure. Similarly, Grieffenstien and colleagues (1995) found that, in a sample of brain injury patients referred for personal injury neuropsychological evaluation, scores on PVTs and SVTs were not significantly related. Based on factor analyses which were not in support of a unitary construct of malingering, the authors of this study concluded that the evaluation of malingering should be approached as a multifaceted construct consisting of both performance and symptom-endorsement factors, that each contribute unique information. These findings support recommendations to use a variety of measures, which tap different constructs via different methods, when conducting a comprehensive evaluation of psychological injury.

Diagnostic Validity

Diagnostic validity or utility, in contrast to construct validity, refers to a test's usefulness in differentiating between individuals with a condition (or a diagnosis) or without the condition (Larrabee & Berry, 2007). In the context of malingering, diagnostic validity refers to the ability of a test to accurately identify those who are malingering or feigning symptoms versus those who are responding honestly. Diagnostic validity is evaluated based on classification accuracy statistics, such as specificity, sensitivity, positive predictive power, and negative predictive power, which allow researchers to determine the number of subjects in a given sample who were identified as malingering, as a function of the base rate of malingering. Base rate is the relative frequency of a condition or attribute within a given population, or the underlying estimated prevalence of malingering within a particular context. Specificity (SP; true negatives) is the probability of finding a negative (i.e., not malingering) test result for subjects who are indeed not malingering. Alternatively, sensitivity (SE; true positives) is the probability of a positive test result (i.e. malingering) for subjects who are indeed malingering. Using base rate information, researchers can calculate positive predictive power (PPP) and negative predictive power (NPP), which refer to the probability of accurately classifying a subject as either malingering or honest responding, based on a positive or negative test result (Wisdom, Callahan, & Shaw, 2010).

Receiver Operating Characteristics

Receiver Operating Characteristics (ROC) analysis is a statistical method used to provide information about the classification accuracy of a measure and to allow clinicians to select cut-scores which will optimize both specificity and sensitivity. As noted by Pintea and Moldovan (2009), ROC analysis is useful when researchers seek to use a test to discriminate between two groups, aim to determine optimal cut-off points, and wish to compare the performances of two or more tests. ROC analysis produces visual representations of sensitivity plotted against 1-specificity for a given measure across two groups. These graphical representations, known as ROC curves, provide valuable information through both elevation and curvature. Measures which are determined to be useful classifiers will have an initial steep incline (indicative of when the true positive rate is high while the false positive rate is low), followed by a tapered curve and then a levelling off. The levelled area is indicative of the point at which the false positive rate increases rapidly while the true positive rate stalls. When sensitivity is equal to the false positive rate, the resulting graph is a diagonal line reflecting the performance of a truly random test (Pintea & Moldovan, 2009). In these cases, the measure's classification ability to discriminate between the two groups is no better than chance. Lines that depart from the diagonal line are indicative of a measure that is capable of discriminating between the two groups (i.e., it has the capacity to identify true-positives while reducing the number of false-positives; Fawcett, 2006).

To compare across classifiers, researchers use a statistic known as the Area Under the Curve (AUC). The AUC is interpreted as the probability that an individual who is randomly selected from the target group (as defined by the researcher) will have a higher score on the measure of interest than a randomly-selected individual from the non-target group (Pintea & Moldovan, 2009). The AUC value will always range from 0 and 1, with higher values indicating greater probability of correct classification, as an AUC value of .5 is indicative of a classification capability that is no greater than chance (i.e., 50%). AUCs between 0.5 and 0.7 are suggestive of low diagnostic or classification accuracy, with AUCs ranging from 0.7 to 0.9 being considered moderately accurate (Pintea & Moldovan, 2009). In addition to providing information about the classification abilities of a measure, the value of the AUC is equally informative when not significant, as this suggests that the measure is a poor classifier of the intended measured construct or that the measure may be a poor screener in the given context.

ROC analysis additionally allows for the evaluation of the classification properties of a measure at various cut-scores, which is valuable in clinical settings for determining at what value specificity (true negatives) and sensitivity (true positives) are optimized for a screening measure in predicting a dichotomized outcome (e.g., feigning vs. not-feigning; Greiner, Pfeiffer, & Smith, 2000). An optimal cut-point is the score at which detection of true positives and true negatives is maximized (and therefore false-positives are minimized). ROC analyses provide information on the sensitivity, specificity, positive predictive power, and negative predictive power for each possible cut-point value. ROC analysis additionally allows for the evaluation and comparison of true positive rate (the percentage of cases correctly classified) and the false positive rate (the percentage of cases incorrectly classified) as the set criterion, or cut-point varies (Fawcett, 2006). In sum, this information can be used to determine optimal cut-points for various contexts in which the measure is being employed. As noted by Larrabee and Berry (2007), when evaluating the potential for malingering, most researchers and clinicians set high specificity levels in order to minimize the possibility of false-positive errors, or the misidentification of a honest responder as a malingering. As a consequence, the potential for true malingerers to go undetected increases. Authors emphasize that this trade-off, in which cut-scores are chosen that maximize specificity at the expense of sensitivity, thereby allows tests to be valuable at “ruling-in” malingering, but deficient at ruling it out (Larrabee & Berry, 2007).

ROC Analyses for Symptom and Performance Validity Tests

Miller (2001) conducted ROC analyses using M-FAST total scores during the initial measure validation process, using both clinical (psychiatric patients) and non-clinical samples (undergraduate students). The clinical sample, which consisted of 50 psychiatric inpatients, was divided and sorted into malingering and honest responding groups based on scores on the SIRS, and this method of differentiation was validated using independent samples t-tests to verify that M-FAST total scores were significantly different across the groups. The non-clinical sample was comprised of 116 undergraduate students, who were instructed to either simulate mental illness in a “realistic manner” or to respond to interview questions honestly. According to study protocol, students in the simulation group were informed that persons will attempt to fool mental health professionals (by faking illness) in order to avoid incarceration, to gain money from a lawsuit, or for financial compensation from employers or the government. They were additionally informed that there are methods of detecting feigned mental illness. When using a clinical sample, the AUC was reported as .95 (SE = .03). Authors report this as a significant elevation above chance. For the non-clinical sample, AUC was reported at .99 (SE = .01), which is also indicative of a significant elevation above chance. Results of these ROC curve analysis are suggestive that the M-FAST has valuable diagnostic ability to identify malingering, with both the clinical and non-clinical samples. Utility analyses, conducted on both the clinical and non-clinical samples, identified a cut-off score of 6 as "highly suggestive" malingered symptomatology, and suggestive that further testing is necessary. Authors note that this cut-score is intended to minimize false negatives (or the percentage of actual malingerers who are not identified).

As noted by Christiansen and Vincent (2012), the M-FAST was originally developed for the identification of feigned psychotic disorders, as the majority of questions are limited to presumed psychotic symptoms. As such, it holds questionable utility for the detection of other forms of feigned mental illness, such as PTSD, and, by extension, questionable utility for situations in which feigning of non-psychotic type illnesses is more likely, such as civil litigation. Researchers have attempted to address these concerns, using coaching paradigms and known-groups comparison designs within the context of civil litigation (Guriel-Tennant & Fremouw, 2006; Alwes, Clark, Berry, & Granacher, 2008); however, notable limitations in these studies call for continued investigation. In particular, as outlined by Christiansen and Vincent (2012), past studies have suffered from a tendency to assume either total honesty or total deceit, which is likely incongruent with real-world situations. Likewise, studies on distinguishing feigned cognitive impairment in an outpatient civil forensic setting have demonstrated difficulty parsing out individuals identified as malingering who report only psychiatric symptoms, absent of brain injury or other neurological insults. Lastly, use of the manualized cut-off score of 6 for the M-FAST, which has been common practice in both research and clinical applications, has not been thoroughly validated in civil litigation settings.

In addressing some of the methodological limitations noted above, Christiansen and Vincent (2012) utilized a between-subjects simulation design to evaluate the utility of the TSI-ATR and the M-FAST in detecting simulated post-traumatic stress symptoms in a simulated personal injury paradigm. Participants were randomly assigned into four groups, including an active litigation with suggestion to exaggerate symptoms and a no-litigation condition. Using ROC analysis, they found that both the M-FAST total score and the TSI ATR subscale deviated from the line of no information (AUCs M-FAST = .65; TSI ATR = .64), which is suggestive that the measures have some discriminative ability; however, neither held significantly better predictive utility than the other. By entering both into a single linear model, it was found that both measures contributed unique information regarding participant response styles; however, the full model did not yield significantly better predictive ability over either individual measure alone. Within this paradigm, a recommended cut-score of 6 for the M-FAST (raw score) yielded a sensitivity of .54, and a recommended cut-score of 61 for the TSI ATR (T-score) yielded a sensitivity of .42, meaning that there is a 54% and 42% probability, respectively, that a person scoring higher than these cut-scores is malingering.

Researchers Widows and Smith (2005) have also used ROC curve analysis to evaluate the ability of the SIMS total score to discriminate between a pooled sample of 298 honest responders (including undergraduates and psychiatric patients) and instructed malingerers, who were asked to simulate either amnesia, schizophrenia, or neurological problems (Merckelbach & Smith, 2003). The resulting ROC curve was found to have good overall performance in discriminating between those asked to simulate various psychiatric conditions or those responding honestly (AUC = .96; SE = .002).

According to a recent meta-analysis (van Impelen, Merckelbach, Jelicic, & Merten, 2014), there have been 24 identified simulation studies utilizing the SIMS. The majority of these studies have used undergraduate samples, who were instructed to feign psychopathology (either psychiatric or cognitive symptoms) or serve as honest-responders. As noted by van Impelen and colleagues, a majority of these studies relied on subject's lay-person understanding of psychological disorders; however, some utilized subjects who may have a greater understanding of psychological symptoms past on past experience, or who were instructed or coached regarding specific symptoms. The researchers of this meta-analysis concluded generally that the SIMS is effective in differentiating between simulated feigners and honest responders, highlighting effect sizes ranging from .5 to 4.7. They note that differences in control groups (e.g., either true patient populations or non-clinical controls) may have contributed to the variability in effect sizes. In the non-clinical control groups, specificity rates ranged from .88 to 1.00 (when using the cut-scores of >14 or >16, respectively). It is noted that specificity rates for clinical control groups ranged from .23 to .83, sensitivity rates ranged from .87 to 1.00 for "naïve" or un-coached groups. Researchers in this study highlighted, also, the importance of coaching and prior psychological knowledge, noting that while coaching (or warning of validity indicators) undermines sensitivity, while prior knowledge of psychopathology does not. A counterintuitive relationship was identified, finding that prior knowledge of psychopathology significantly reduced the degree to which sensitivity was undermined by coaching or warning of validity indicators. Researchers additionally note how the SIMS operates by identifying over-reporting of psychopathology, which is different from PVT assessments, which focus on underperformance. Despite this difference in approaches, significantly elevated SIMS scores have been found in groups asked to simulate neurocognitive deficits. Comparative sensitivity rates have been found between the SIMS and TOMM in studies involving individuals coached to feign cognitive problems (SE = .87 and .86, respectively). The authors conclude that the SIMS "cannot be relied upon" when attempting to detect feigned cognitive impairment, which calls into question its general utility in personal injury and PTSD cases, where cognitive symptoms may be present.

Gaps in the Literature

Based on the above discussion and literature review, there are several areas which warrant further investigation. First, researchers have emphasized the importance of utilizing several malingering measures in efforts to lend incremental validity to conclusions and to rule out the possibility of false positives (Larrabee & Berry, 2007). Likewise, researchers have highlighted the need to conceptualize malingering as a non-unitary, dynamic construct (Greiffenstein, Gola, & Baker, 1995), which is comprised of various facets, including both performance and symptom-endorsement. Furthermore, in generating evidence in support of construct validity, the importance of evaluating the impact of method variance has been noted. Researchers have reported the extent to which various measures of malingering assessment are correlated as evidence for convergent and discriminant validity; however, to the author's knowledge, few researchers have evaluated construct validity of specific measures of malingering by utilizing a MTMM matrix. Similarly, there exists a need for further validation research designed to evaluate and compare the classification capabilities of various PVTs and SVTs in various medico-legal contexts.

Secondly, while several researchers have attempted to evaluate the classification accuracy of various malingering measures using ROC curve analysis, few researchers have utilized a personal injury litigation paradigm, which differs from commonly used criminal adjudication paradigms. Given the relative frequency of civil litigation cases, in comparison with criminal cases, further research investigating the context-specific impact of civil litigation is warranted. When using personal injury paradigms, researchers have focused either on coaching effects or the identification of malingered neurocognitive deficits. While valuable, there remains a need to understand the impact of factors such as suggestion to malinger (distinct from coaching or a directive), and the isolated effect of psychological impairment without co-occurring brain injury.

Purpose of the Present Study

Due to the estimated prevalence of feigning in personal injury cases, the need for valid and interpretable measures of malingering that are reflective of the context-specific features of feigned or malingered psychopathology is substantial. The legal, financial, and psycho-social consequences inherent in a litigation context necessitate the need for a clear understanding of (a) how various response styles present with a medico-legal paradigm and (b) the reliability, validity, and usefulness of methods available to detect dishonest responding. While there exists a substantial body of research which addresses the evaluation of malingered brain-injury and neurocognitive deficits in personal injury cases, comparatively little research has addressed the importance of feigned psychiatric and neurocognitive deficits outside the context of brain-injury. Likewise, much research has been devoted to the detection of feigned psychopathology in criminal adjudication cases or cases involving feigned psychotic disorders, while little research has addressed civil litigation. Despite prevalent recommendations to utilize multiple measures of malingering within a larger assessment battery, little is known regarding how various measures operate interdependently and the relative value of using measures that address different sub-constructs (e.g., feigned memory deficits), via different methods (in-person interview vs. self-report form).

The present study aims to expand upon previous research by further exploring the relationship between performance validity measures, such as the Test of Memory Malingering, and symptom validity measures, such as the Trauma Symptom Inventory, in the detection of malingering within a simulated personal injury paradigm. Current practice in the assessment of malingered symptomatology when PTSD is alleged supports the use of performance validity measures in addition to traditional symptom validity measures; however, the inter-relationships between various measures of malingering used in this context have not been systematically evaluated. Researchers in the present study aim to evaluate the extent to which symptom-based measures of malingering (SVT) versus performance-based or effort-based measures of malingering (PVT) are correlated, in efforts to lend support for convergent and discriminant validity utilizing a MTMM matrix.

Additionally, the extent to which individual performance and symptom validity measures allow for the determination of malingering was examined using ROC curve analyses. Several researchers have highlighted the need for continued evaluation of diagnostic classification abilities of malingering measures, noting the context-specific nature of previously reported cut-scores and utility statistics. Previous investigations of the diagnostic utility of the M-FAST and TSI in a personal injury simulation paradigm found significant, but low classification accuracy. The present study aims to expand upon this previous research, utilizing the revised TSI-2-A (Alternate version; henceforth referred to simply as the TSI-2) and additional prominent measures of malingering (the TOMM and SIMS).

The present study used a between-groups simulation design, in which participants were randomly assigned to one of four experimental conditions, as detailed through a written vignette (see Appendix B). Each vignette indicated that the participant had recently been involved in motor vehicle accident and was continuing to experience a series of psychological and cognitive symptoms associated with the accident. In addition, participants were informed that they were either: (1) asked to complete an evaluation at the request of his or her physician following the conclusion of a lawsuit (Condition 1: post-litigation, no suggestion), (2) simply asked to complete an evaluation at the request of his or her physician (Condition 2: no litigation and no suggestion), (3) asked to complete an evaluation at the request of his or her attorney for the purposes of an ongoing case (Condition 3: active litigation, no suggestion); or (4) asked to complete an evaluation at the request of his or her attorney for the purposes of an ongoing case, with the suggestion, by the attorney, that greater impairment would lead to a larger monetary award (Condition 4: active litigation and suggestion). Under this paradigm, participants were then instructed to respond to a set of measures as if they were the person in the vignette they received. Measures included the TSI-2, M-FAST, TOMM, and SIMS, which were administered in person (TOMM; M-FAST) and via computer (SIMS; TSI-2).

The following hypotheses are proposed:

H1: Correlational analyses will be conducted across all bivariate combinations. According to MTMM matrix predictions, participant scores will correlate according to measurement of underlying traits and methods, such that measures psychological symptom endorsement (TSI-2 ATR, SIMS, M-FAST) will correlate more highly with other measures of symptom endorsement than scores on measures of performance or effort (i.e., TOMM). Likewise, it is predicted that measures associated with cognitive symptoms (TOMM; SIMS NI and AM) will correlate more highly than with measures of affective experience. Furthermore, it is predicted that measures will correlate based on similar method, such that in-person measures (M-FAST and TOMM) will be more highly correlated than with measures completed on the computer.

H2A: Using the dichotomous classifier of instructional condition (conditions 2 vs. 4), ROC analysis will be conducted to examine the diagnostic utility of PVT (TOMM) and SVT measures (M-FAST, SIMS, TSI-2 ATR) for detecting malingering in a personal injury litigant population feigning PTSD. A range of cut-off scores will be examined for each measure.

H2B: Using groups dichotomously classified as malingering based on SIMS total score, ROC analysis will be conducted to examine the diagnostic utility of PVT (TOMM) and SVT measures (M-FAST, TSI-2 ATR) for detecting malingering in a personal injury litigant population feigning PTSD. A range of cut-off scores will be examined for each measure.

Method

Participants

This study was a part of a larger research initiative in which graduate researchers and research assistants recruited participants, who were enrolled in psychology courses at a large, public, Southwestern university. All students were recruited through the university's online recruitment pool, SONA, through a study entitled "Motor Vehicle Accident Study." Students received six hours of experiment participation credit for their involvement in this study, which required approximately 3 hours of in-person assessment time.

Inclusion & Exclusion Criteria.

Inclusion criteria required that participants be over the age of 18, enrolled as an undergraduate student at the university, and proficient in English. Participants were screened for inclusion through the SONA website.

Exclusion criteria included those who indicate that they do not hold a valid driver's license, in order to maintain the presumption that participants could reasonably assume the role of the victim of a motor vehicle accident situation similar to those described in the experimental condition vignettes. The initial sample consisted of 458 completed protocols, of which 34 cases were excluded for not holding a valid driver's license. 13 cases were excluded due to errors related inattentive responding. After application of the exclusionary criteria, the final full sample included 411 undergraduate students.

Preliminary descriptive characteristics revealed that participant ages ranged from 18 to 58 (M = 23.04; SD = 5.59), with 68.9% being female. Participants classified themselves as one of the following ethnic or racial categories: Caucasian (22.9%), African American (16.1%), Hispanic (28%), Asian American (23.1%), or Other/Multiracial (10%). Over half of the sample indicated that they are majoring in Psychology, and 69.8% indicated that they had attained an educational standing of Junior or higher. A third of students indicated previous involvement in a motor vehicle accident, and 18% percent indicated previous involvement, either personally or through family, in a civil litigation case. Approximately 15% percent of participants indicated having undergone previous psychological assessment outside the context of a research study.

Materials

Pre-Questionnaire.

A 52-item questionnaire was developed for the current study and was used to collect demographic and background information about the participants (Appendix A). Demographic and background information included variables such as age, gender, marital status, ethnic identification, education level and college major, occupational status, voter registration status, and history of military involvement. The pre-questionnaire also included questions regarding past litigation experience, either as a litigant or through relation to someone involved in a lawsuit. Additionally, participants were asked to answer questions regarding their and their relatives' past motor vehicle accident experiences in which they may have sustained physical or psychological injuries and subsequent legal outcomes, if any. Lastly, participants were asked to detail their involvement and experience with the mental health profession, including past personal history of psychiatric distress and treatment.

Instructional Conditions/ Case Scenario.

The instructional conditions (detailed in Appendix B) under which participants responded included asking participants to imagine that they had been in a motor vehicle accident with a truck owned by a major retail wholesaler. Each vignette details that the participant did not sustain any serious physical injuries, but that they were still experiencing some emotional difficulties as a result of the accident. Symptoms detailed included: jumpiness/nervousness while driving, avoidance of the location of the accident, avoiding conversations about the accident, having bad dreams about the accident, and having an exaggerated startle response. Participants were informed that they were either: (1) asked to complete an evaluation at the request of his or her physician following the conclusion of a lawsuit (post-litigation, no suggestion), (2) asked to complete an evaluation at the request of his or her physician (no litigation and no suggestion), (3) asked to complete an evaluation at the request of his or her attorney for the purposes of an ongoing case (active litigation, no suggestion); or (4) asked to complete an evaluation at the request of his or her attorney for the purposes of an ongoing case, with the implication, by the attorney, that greater impairment will lead to a larger monetary award (active litigation and suggestion).

Post-Questionnaire.

Following completion of the primary measures in the study, participants were presented with a 15-item post-questionnaire which aimed to capture their perceptions of the instructional conditions and the extent to which they felt they were faithfully able to respond to the presented measures as if they were the person in the given scenario (Appendix C). Additionally, questions included a brief verification of participants understanding of specific terms in civil litigation (e.g., “pecuniary damages”), and questions related to the individual’s perception of reasonable amounts of damage awards for this case. The post-questionnaire included two specific items which can be used to gauge condition adherence and may serve as manipulation check. Specifically, participants were asked whether, in the condition that they received, there was the suggestion that greater injuries would lead to more compensation and for what purpose was the evaluation conducted (either physician's request or for litigation purposes). Responses on these items may be used to determine how well the instructional conditions were interpreted and recalled throughout the experiment.

Measures.

Trauma Symptom Inventory – 2 – Alternative (TSI-2-A; Briere, 2011). The TSI-2-A is a 126-item, broadband self-report rating scale of trauma-related symptoms and behaviors. It was designed for use with individuals 18 years and older, and requires a fifth-grade reading level. The TSI-2-A is an alternative form of the larger TSI-2, which consists of 136 items and includes questions specific to sexual trauma. Like the original TSI, the TSI-2 is applicable for use in a variety of inpatient, outpatient, and community settings, and was validated on both men and women in the general population. Items consist of a variety of cognitive, affective, and physiological symptoms in addition to behaviors and experiences commonly associated with trauma. Participants are asked to indicate how often they have experienced these symptoms, such that 0 = *has not happened at all*; 1 = *has happened in the last 6 months, but only rarely*; 2 = *has happened sometimes in the last 6 months*; and 3 = *has happened often in the last 6 months*. Items of the TSI-2 and TSI-2-A, with the exception of sexual trauma items, are the same. For simplicity, the present study will refer to both as the TSI-2. The TSI-2 yields two validity scales (RL and ATR), four factors (Self-Disturbance, Posttraumatic Stress, Externalization, Somatization), and 12 clinical scales and subscales. The TSI-2 RL validity scale assesses a respondent's tendency to deny commonly endorsed symptoms, whereas the TSI-2 ATR scale addresses respondent's tendency to over-endorse trauma symptoms, even when compared to those with confirmed, severe post-traumatic symptomatology. A high score on the TSI-2 ATR scale can be reflective of a variety of phenomena, including: generalized over-endorsement across all items, specific over-endorsement on items associated with PTSD, random responding with over-endorsement on clinically rare symptoms, or very significant levels of distress. As recommended by Briere (2011), a raw score of 15 or higher on the ATR scale indicates invalidity and serves as a cut-off in clinical and forensic settings. In research settings, a cut-off score of 8 is recommended (Gray, Elhai, & Briere, 2010).

According to Briere (2011), the TSI-2 features considerable revision to the original TSI ATR subscale. Researchers Gray, Elhai, and Briere (2010) conducted a discriminant function analysis using a sample of 124 university students, divided into a simulation group and an honest responding group. Honest responders were determined based on met criteria for PTSD using the PTSD Checklist (PCL). Analysis indicated that the TSI-2 ATR had good predictive validity (sensitivity = .75 and specificity = .74) for detecting malingered PTSD. According to Briere (2011), the TSI-2 ATR's predictive ability was "markedly superior" to the original TSI ATR scale. Recommendations to use cut-off scores of 8 (for a sensitivity of .65 and a specificity of .83) in research samples were based on this study.

Test of Memory Malingering (TOMM; Tombaugh, 1996). The TOMM is a three-part testing battery which consists of 50 items per trial. It is used to assess exaggeration and fabrication of memory impairment, and is used as a proxy for evaluation of effort. In Trials 1 and 2, subjects are presented with a series of 50 line drawings of ordinary objects for 3 seconds at a time. After the initial learning phase, where subjects are told to commit the drawings to memory, subjects are presented with 50 sets of pairs of items, in which one item in each pair is identical to a drawing previously presented during the learning phase. Subjects are asked to identify which of the two objects in each set was presented previously. Subjects who are unable to correctly identify 45 of the 50 items presented on Trial 2 are given a retention trial after approximately 15 minutes. The TOMM has been validated on several populations, including neurological patients, college student normal controls and simulators, persons feigning traumatic brain injury and true brain injury litigants, persons with depression, and elderly patients. Scores less than the designated cut-off of 45 on any trial call into question the validity of the test-taker's overall performance. Using the detection strategy of a floor-effect, the TOMM is often perceived as being more difficulty than it truly is. Individuals who are attempting to exaggerate symptoms often perform worse than those with genuine impairment.

Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005; Smith & Burger, 1997). The SIMS is a 75-item self-report measure developed to serve as a screening instrument for the detection of exaggerated or feigned psychopathology and cognitive dysfunction. It was designed for use with individuals aged 18 years and older across a variety of clinical and forensic settings. The SIMS is a domain specific, multi-scale measure which yields both a total score that reflects general feigned presentations, as well as five non-overlapping subscales including: psychosis (P), neurologic impairment (NI), amnestic disorders (AM), low intelligence (LI), and affective disorders (AF; Widows & Smith, 2005). The SIMS has been validated on a wide-range of clinical, forensic, and community samples. As noted by the authors, the SIMS is not intended to be a determinant of feigned presentations in isolation. Rather, it is recommended that the SIMS be used in conjunction with other measures, including structured interviews and performance-based tests in order to provide convergent data for the classification of feigning. The SIMS total score is interpreted as an overall estimate of the likelihood that the subject is feigning or exaggerating psychological symptoms or cognitive impairment. The total score has demonstrated adequate utility for the identification of potential feigned response styles. Recommended cut-off scores for the SIMS were based on a sample of 238 participants who were either asked to respond honestly or asked to simulate one of six experimental conditions, including simulated psychosis, amnesia, depression, low intelligence, neurologic impairment, and faking bad. Cut-off scores were identified by maximizing the "hit rate," or percentage of correctly classified cases, for each group, when comparing each simulation group to the honest responding group (Lezak, 1995). Using the Total score, a cut-off of >14 yielded a sensitivity of 95% and a specificity of 88%. A cross-validation study (Smith & Burger, 1997) supported the use of the SIMS Total score as the most effective indicator to differentiate between honest and malingering respondents.

Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001). The M-FAST is a 25-item structured interview designed as a screening measure for the determination of malingered psychopathology. It is a widely used forensic assessment measure intended to assess response style. The M-FAST includes seven subscales based on empirically validated response styles and reporting strategies used by malingering individuals. The subscales measure: discrepancies between observed and reported symptoms (RO), extreme symptomatology (ES), rare combinations of symptoms (RC), unusual hallucinations (UH), unusual symptom course (USC), overly negative self-image (NI), and suggestibility (S). While four subscales (UH, RC, RO, and ES) have been reported as consistently able to discriminate between honest responders, known malingerers, and those instructed to malinger, the total score is the most effective and most frequently used. According to Miller (2001), a cut-off score of 6 is considered the most appropriate for differentiating between true and false responding; however, it is suggested that this cut-off may be inappropriate for personal injury litigation populations (Christiansen & Vincent, 2012). The M-FAST has been standardized and was validated using both known-groups clinical samples and simulation non-clinical samples. Inter-rater reliability is reported to be better than 99% and total score reliability estimates of .93 for clinical samples and .92 for non-clinical samples have been reported. Analyses of criterion, convergent and discriminant validity have been conducted using the SIRS, M-Test, and MMPI-2, and have found to be acceptable.

Design and Procedure

Recruitment.

Participants were recruited via announcements made by undergraduate research assistants in their undergraduate courses and via the university's online psychology research recruitment pool, SONA. Participants were informed of the benefits of their participation, which included earning extra credit hours which can be allotted to the psychology course of their choice. Participants were required to sign up and were granted credit via the SONA system. Via an announcement on the SONA website, potential participants were informed that they would qualify for the study if they were 18 years of age, currently enrolled in an undergraduate course, and are proficient in English. English proficiency was required due to the fact that the measures administered in this study were only done so in English.

Participation Phase.

Participants completed the research protocol at individual computer stations in an undergraduate research space during scheduled sessions. All self-report measures, in addition to the pre- and post-questionnaires, were administered via the online survey platform, Qualtrics. Participants were assured of the anonymity of their responses prior to receiving information about the study, and in-person assessments (i.e., the TOMM and M-FAST) were administered in a private office space. After granting informed consent, participants were asked to complete the pre-questionnaire, LEQ, TEC, and IES. Participants then read one of four instructional scenarios (detailed in Appendix B), which were randomly assigned using Qualtrics. Research assistants, who administered the in-person assessments and supervised the completion of the online administration of the self-report measures, were blinded to the instructional conditions of the participants. After reading the assigned vignettes, participants completed the M-FAST, SIMS, TOMM, Brief Symptom Inventory (BSI), Shipley-2, and TSI-2. These measures were presented in a random order by way of the Qualtrics randomization feature. Following these measures, participants completed the post-questionnaire. Participants were debriefed as to the nature of the study, granted participation credit, and excused.

Data Analytic Plan

Before addressing the specific aims and hypotheses, descriptive data regarding participant demographic information, litigation history, and mental health history were generated. Additionally, descriptive testing data including means, standard deviations, and failure rates, across the total sample and individual conditions, were generated. Analyses were conducted to identify and account for significant differences amongst instructional conditions based on age, education, race, and gender.

All participants were examined with the following self-report measures of malingered symptomatology and in-person tests of malingered cognitive functioning: the TOMM (Tombaugh, 1996); M-FAST (Miller, 2001); SIMS (Widows & Smith, 2005); and TSI-2 (Briere J. , 2011). The TOMM consists of two trials, with an optional third retention trial dependent upon performance on Trial 2. For the purpose of these analyses, the results of Trial 2 of the TOMM was used. Scores on the TOMM Trials 1 and 2 range from 0 to 50, with higher scores indicating optimal effort or normative performance, as the TOMM operates using the floor effect. Scores on the M-FAST range from 0 to 25, with greater scores indicating greater likelihood of a feigned presentation. Subscale scores on the TSI-2 ATR range from 0 to 24, with higher scores indicating a greater likelihood of an invalidated clinical profile due to excessive symptom endorsement. On the SIMS, subscale scores range from 0 to 15, with greater scores serving as indicative of elevated endorsement of symptoms consistent with each of the five subscale domains. The SIMS Total score ranges from 0 to 75, with higher scores reflecting a greater endorsement of atypical, improbable, or inconsistent symptoms. Data analysis was conducted using SPSS statistical software.

Using a between-subjects simulation design, the present study first aims to evaluate bivariate correlations amongst four prominent measures of malingering: TOMM, M-FAST, SIMS, and TSI-2. Consistent with assumptions of the MTMM, participant scores were expected correlate according to measurement of underlying traits and methods. Convergent validity would be demonstrated through statistically significant correlations between measures related to psychological symptom endorsement (TSI-2 ATR, SIMS, M-FAST total scores). Divergent validity would be demonstrated through weaker correlations between scores on measures of psychological symptom endorsement and scores on measures of performance or effort (i.e., TOMM). Lastly, examination of correlations between measures which utilize similar methodologies would provide evidence of method variance. It is hypothesized that in-person measures (M-FAST and TOMM) would be more highly correlated than with measures completed on the computer.

Secondly, the present study aimed to partially replicate and extend findings from Christiansen & Vincent (2012) in supporting the predictive utility of the M-FAST and TSI ATR subscale as valuable classifiers of malingered protocols. Specifically, using ROC curve analysis, the present study aims to evaluate the abilities of the M-FAST, TSI-2 ATR, SIMS, and TOMM to discriminate between those told to “simulate” (i.e., endorse) specific symptoms from those told to “malinger” (i.e., endorse and exaggerate) symptoms. We predicted that the M-FAST total score, SIMS total score, TOMM trial 2 score, and TSI-2 ATR would yield significant discriminant ability in predicting group membership, based on previous research. Two hypotheses have been proposed:

For Hypothesis H2A, ROC analysis was conducted to examine the diagnostic utility of the TOMM, M-FAST, SIMS, and TSI-2 ATR. Because ROC analysis requires the use of a binary or dichotomous classifier, only individuals who were randomly assigned to Conditions 2 and 4 were used in this first analysis. The use of Conditions 2 and 4, which refer to the “no-litigation” and “litigation with suggestion to malinger” groups, approximated honest-responding and malingering, respectively. While it is noted that it is unreasonable to assume that all individuals in these conditions were able to faithfully subscribe to the instructions of the condition throughout the research protocol, the likely variability found in the “malingering” and “honest responding” groups is ecologically valid, as there is expected variability in plaintiff’s responding in true litigation settings. ROC analyses were used to examine the utility of the four malingering measures by plotting the sensitivity against 1-specificity. Using these plots, diagnostic classification statistics were calculated and optimal cut-points derived for each measure. Using this information, the measures will be evaluated and compared for their respective diagnostic utility in detecting malingering in a personal injury context.

The data analytic procedure for H2B will mirror that of H2A, with the exception that the groups were dichotomized based on SIMS total score. ROC curve analyses are traditionally conducted using an external or criterion classifier, which is used to create comparison groups. The present study proposes to use the SIMS total score as a means of classifying individuals as either “probable honest responders” or “probable malingerers.” The SIMS serves as an appropriate classifier, compared to the other three measures, based on its broader inclusion of possible feigned psychopathologies, instead of a narrower focus on memory impairment (TOMM), psychosis (M-FAST), or trauma (TSI-2). Individuals scoring higher than the recommended cut-score (greater than 14) on the SIMS were placed in the “probable malingerers” group. ROC analyses will then examine the utility of the remaining malingering measures (TOMM, TSI-2 ATR, and M-FAST) using the method proposed above.

Results

Sample Characteristics and Preliminary Analyses

As a part of a larger research initiative, the full sample consisted of 458 completed protocols. Thirty-four cases were excluded for not holding a valid driver’s license. Thirteen cases were excluded due to errors related inattentive responding. After application of the exclusionary criteria, the final full sample included 411 undergraduate students.

Demographics.

Out of 411 protocols including in the final analysis, 283 participants self-identified as being female (68.9%) and 128 participants identified themselves as being male (31.2%).

Participant ages ranged from 18 to 58 (M = 23.04; SD = 5.59). The distribution of ages is notably skewed, with a substantial portion of participants being 25 years of age or younger, as would be expected in an undergraduate sample. Participants identified themselves as belonging to one of the following ethnic or racial categories: Caucasian (22.9%), African American (16.1%), Hispanic (28%), Asian American (23.1%), or Other/Multiracial (10%). Over half of the sample indicated that they are majoring in Psychology, and 69.8% indicated an educational standing of Junior or higher.

Of the total analytical sample, approximately one third of students (33.3%) indicated previous personal involvement in a motor vehicle accident, in which the participant sustained physical or psychological injuries. Most participants indicated having settled with the other driver's insurance company as a result of these experiences (16.8%), with less than 10% of participants indicating that they neither sought nor received compensation. Eight percent of participants indicated having previous involvement in a motor vehicle accident in which the participant was at fault, and inflicted physical or psychological injuries on another. Again, in these instances, most participants indicated that the incident was settled via an insurance company. Over 60% of participants indicated having a close relation who was involved in a motor vehicle accident in which psychological or physical injuries were sustained.

Eighteen percent of the total sample (n = 74) indicated previous involvement, either personally or by means of a family member, in a lawsuit, either criminal or civil. Of those participants who provided sufficient detail regarding the types of previous lawsuits in which they were involved, approximately 36% (n = 27) indicated involvement in a civil suit, with 22% of those cases involving a motor vehicle accident.

Approximately a third (30.2%) of all participants indicated having previously received mental health treatment, and approximately 15% indicated having previously undergone a psychological evaluation outside the context of research. Seventeen percent of the sample indicated having received a past psychological or psychiatric diagnosis. According to self-reported diagnoses, captured via qualitative response options, 57% of those with previous diagnoses experienced depression; 34% experienced an anxiety disorder; 20% experienced an attention deficit or hyperactivity disorder, and approximately 11% experienced bipolar disorder. Less than 10% of those with self-reported previous diagnoses indicated having been diagnosed with PTSD, OCD, or a personality disorder. Forty-one percent of those with previous diagnoses indicated having been diagnosed with one or more conditions.

Condition Assignment.

Participants were randomly assigned into one of the following four conditions: Condition 1: Post-litigation, No suggestion (n = 105; 25.5%); Condition 2: No litigation, No suggestion (n = 104; 25.3%); Condition 3: Active litigation, No suggestion (n = 106; 25.8%); Condition 4: Active litigation and Suggestion (n = 96; 23.4%).

There were no significant differences across conditions based on participant age [F (3,407) = 1.501, p = .212], gender [$X^2(3)$ = 2.97, p = .397], race/ethnicity [X^2 (12) = 12.69, p = .392], or level of education [X^2 (12) = 8.59, p = .737]. Conditions also did not significantly differ in terms of personal motor vehicle accident history [X^2 (3) = .60, p = .897]. A summary of descriptive statistics by condition is presented in Table 1.

Testing Statistics.

Descriptive test statistics, including means, standard deviations, and failure rates for the TSI-2 ATR, SIMS, M-FAST, and TOMM Trial 2 (T2), across the full sample, are presented in Table 2. Testing statistics per condition are also presented. Across the full sample, total scores for the TSI-2, SIMS, and M-FAST were normally distributed, with skewness of 1.02, .86, and 1.08 (SE = .12; respectively) and kurtosis of .11, .35, and .69 (SE = .24). Scores for the TOMM Trial 2 were non-normally distributed, with skewness of -4.16 (SE = .12) and kurtosis of 18.05 (SE = .24). The highly negatively skewed distribution of TOMM scores is expected given the test's operation via the floor effect. Transformation of TOMM total scores had no effect on subsequent analyses, and as such, non-transformed data are presented. Histograms of total score frequency across Condition 2 and 4 are presented in Figure 15.

Failure rates are reflective of the percentage of the population whose total score on the given measure fell below the recommended cut-off score for that test. Notably, over 70% of the full sample evidenced failure on the SIMS, having scored a total of 15 points or higher. Based on failure rates, a sizeable proportion of the total sample would be recommended for further evaluation of malingering based on the SIMS and the M-FAST. Failure rates trended in the expected direction, based on instructional condition, with larger failure rates typically found in Conditions 3 and 4.

Manipulation Check.

Before conducting planned ROC analyses, a manipulation check was conducted to determine cross-condition differences between Conditions 2 and 4 on variables of interest. Chi-square tests were utilized to examine cross-condition differences. Using the recommended cut-off scores for each of the malingering measures, it was found that Conditions 2 and 4 did not different significantly on the number of participants classified as "malingering" based on failure of the TOMM T2, SIMS, or M-FAST. Conditions 2 and 4 significantly differed in terms of the number of individuals classified based on failure on the TSI-2 ATR subscale (see Table 4). Based on the results of these preliminary analyses, it was expected that the TOMM T2, SIMS, and M-FAST would not demonstrate significant discriminant ability between Conditions 2 and 4. It was predicted that the TSI-2 ATR subscale would demonstrate significant discriminant ability.

Primary Analyses

Construct Validity.

To address the first hypothesis, bivariate correlations were generated between the four measures of malingering in order to examine convergent and discriminant validity through a multi-trait, multi-method matrix. Results are presented in Table 5. Absolute values of correlations ranged from .15 to .65. It is noted that correlations between the SVTs and the TOMM are negative, which is a product of the fact that lower scores on the TOMM reflect a worse performance. This is reversed from the SVTs where lower scores reflect a more "honest" performance. Therefore, negative correlations between the TOMM and SVTs indicate convergent validity, however weak. As predicted, correlations between the PVT (the TOMM) and the SVTs were weaker than correlations across the three SVTs (ranging from -.15 to .28). Convergent validity between symptom validity measures was demonstrated through moderate, but statistically significant correlations between the SIMS, M-FAST, and TSI-2 ATR. Notably, the correlation between the TSI-2 ATR and the M-FAST (r = .51), while significant, was weaker than correlations between the TSI-2 ATR and the SIMS, or the

SIMS and the M-FAST, suggesting that these two measures may capture different subdomains of the malingering construct.

Diagnostic Validity.

The diagnostic utility of the M-FAST, TSI-2 ATR, SIMS, and TOMM in detecting malingering in a personal injury context was examined through the use of ROC analyses. Two hypotheses were evaluated. First, ROC analyses were conducted on the four measures using instructional conditions 2 and 4 as a dichotomous classifier (H2A). These conditions were selected to serve as proxies for "honest responding" and "malingering" groups. Preliminary comparisons evaluating the effect of the manipulation, or how the groups differentially responded based on instructional condition, suggest that that the M-FAST, SIMS, and TOMM T2 would not demonstrate significant classification ability.

ROC curves, using the binary classifier of instructional condition, are presented in Figures 1-4. All curves deviated from the line of no information in a positive direction. In order to evaluate the significance of each ROC curve departure from the line of no information, areas under the curve (AUCs) were calculated for each measure. As noted above, the AUC value ranges from 0 and 1, with higher values indicating greater probability of correct classification. AUCs between 0.5 and 0.7 suggest low diagnostic or classification accuracy, with AUCs ranging from 0.7 to 0.9 being considered moderately accurate (Pintea & Moldovan, 2009). A summary of AUCs, significance statistics, and 95% confidence intervals for each test are represented in Table 6. Consistent with preliminary analyses, only the TSI-2 ATR produced an AUC that was statistically significant (AUC = .662; *Sig.* = .000). While the TOMM T2 and M-FAST produced curves that deviated from the line of no information, the deviation was not statistically significant from the null model (i.e., from

.500, indicating classification accuracy no greater than chance). The SIMS AUC was not statistically significant using a .05 criterion, however it approached significance. This finding is consistent with preliminary analyses suggesting that the diagnostic utility of the SIMS may be weak. Tables 7-10 detail sensitivity, specificity, positive predictive power, negative predictive power, and classification accuracy rates for each test, for a range of cut-off scores. Calculations are reflective of a base-rate of approximately 50% (48%) as determined by the random assignment into either a “malingering” or “honest responding” condition (Conditions 2 and 4). Optimal cut-points, for this paradigm, are represented in Figures 5-8, and are reflected by the intersection of specificity and sensitivity. Notably, true population base-rates, as noted above, are expected to be much lower than 50%.

Hypothesis H2B also addressed the question of diagnostic validity using ROC curves, however with the use of a criterion classifier. As proposed, groups were dichotomized based on SIMS total score, using the full sample ($N = 411$). Failure on the SIMS (i.e., a total score > 14) served as a proxy categorization as “probable malingerers,” whereas passing the SIMS (i.e., a total score ≤ 14) served to reflect “probable honest responders.” Since the SIMS score will be used as the classifying criterion, only the remaining tests (TSI-2 ATR, M-FAST, and TOMM) will be examined.

ROC curves, using the binary criterion classifier of SIMS pass/failure, are presented in Figures 9-11. All curves deviated from the line of no information in a positive direction. Again, areas under the curve (AUCs) were calculated for each measure. A summary of AUCs, significance statistics, and 95% confidence intervals for each test are represented in Table 11. All measures produced significant AUCs, with the TSI-2 ATR and M-FAST being significantly different from the line of no information at the .01 level. The TOMM T2 also produced a significant AUC, however at the .05 level. Tables 12-14 detail sensitivity, specificity, positive predictive power, negative predictive power, and classification accuracy rates for each test, for a range of cut-off scores. It is noted that based on SIMS total score failure, 71% of the total sample was classified as “probable malingerers.” This is reflective of the use of the SIMS as primarily a screening measure. Calculations are function of this “base-rate” of approximately 71% as determined by the classification as either “probable malingering” or “probable honest responding” condition (as determined by SIMS pass/failure). Optimal cut-points, for this paradigm, are represented in Figures 12-14, and are reflected by the intersection of specificity and sensitivity.

Follow-up Analyses.

During protocol administration, it was noted that several participants verbally indicated to research assistants that they had difficulty remembering to simulate for the duration of the protocol. Additionally, upon subsequent data analysis, it was found that a portion of participants failed post-experimental manipulation checks, suggesting that participants had difficulty retaining information contained in the case vignette or approached the research protocol inattentively or with less-than-optimal motivation. These difficulties would influence the degree to which the groups genuinely responded differently to the tasks, or the impact of the experimental manipulation, thus leading to greater ambiguitiy between groups for the present study. Acknowledging these difficulites, analyses were re-run utilizing a sub-set of the analytical sample where participants successfully answered manipulation checks. By excluding cases based on manipulation check failure, we hoped to improve classification accuracy.

Within the post-questionnaire, participants were asked two debriefing questions:

1. In the scenario, did anyone suggestion that you would receive more compensation if your injuries were greater? (Yes/ No)
2. In the scenario, what was the purpose for which you completed the evaluation? (Physician's request / Litigation purposes)

Based on participant responses, we identified a subsample of 201 cases for which both questions were answered correctly, given the individual's assigned condition. Notably, conditions significantly differed in terms of how many individuals per group failed these manipulation checks, such that participants in Condition 2 evidenced a greater proportion of failures compared to Condition 4 [X^2 (1) = 14.49, p < .001]. Chi-square tests of difference across Conditions 2 and 4, for each of the measures of interest were conducted. Results are presented in Table 15. The above ROC analyses by instructional condition were repeated using this subset of the original sample. Despite approximating a sample of participants for whom the content of the case vignette was retained, classification accuracy was not improved. AUC summaries by malingering measure, utilizing the subsample, are presented in Table 16.

Discussion

The present study expanded upon previous research by further exploring the relationship between performance validity measures and symptom validity measures in the detection of malingering within a simulated personal injury paradigm. As noted previously, current practice in the assessment of malingered symptomatology, when PTSD is alleged, supports the use of performance validity measures in addition to traditional symptom validity measures; however, the inter-relationships between various measures of malingering used in this context have not been systematically evaluated. Given the financial and legal implications associated with successful malingering of PTSD in a civil litigation case, the examination of the relationships between measures of malingering is necessary. The results of the present study demonstrated expected relationships between performance validity and symptom validity measures. As predicted, measures of symptom validity were more strongly correlated with other symptom validity measures and demonstrated weaker relationships to a measure of performance validity (the TOMM). These findings are consistent with previous research on the relationship between PVT and SVT failure as it relates to post-traumatic injury (Greiffenstein et al., 1995; Demakis et al., 2008), suggesting that malingering is a non-unitary construct comprised of both performance and symptom-endorsement factors. Likewise, these findings support clinical recommendations to administer a variety of malingering measures when conducting an assessment, being sure to select measures that will tap into the sub-constructs of performance and symptom validity.

The second aim of this study was to examine the extent to which individual performance and symptom validity measures allow for the detection of malingering. Building on previous research which utilized a simulated personal injury paradigm, the present study examined the utility of prominent measures of malingering in differentiating between honest responders and feigners. This was achieved by using ROC curve analysis, which evaluated each test's ability to appropriately classify responders according to assigned condition and also based on failure on a commonly-used malingering screener. Several researchers have emphasized the need for continued evaluation of diagnostic classification abilities of malingering measures, citing the context-specific nature of previously reported cut-scores and utility statistics. ROC curve analysis has been used historically to support diagnostic validity of a measure by illustrating a tests ability to appropriately classify cases. Diagnostic validity is evaluated based on classification accuracy statistics, such as specificity, sensitivity, positive predictive power, and negative predictive power. These statistics allow researchers to determine how well a test identifies those responses that are consistent with a specific condition (i.e., malingering) and allow for identification of optimal cut-points, based on the intended use of the test.

Using the dichotomous classifier of condition assignment, results suggest that only the TSI-2 ATR is useful in discriminating between the groups, despite all tests deviating from the line of no information. As such, only the TSI-2 ATR evidenced diagnostic classification abilities that were statistically better than chance. Notably, calculations for classification accuracy statistics are reflective of a base-rate of approximately 50% (48%), as determined by the random assignment into either a "malingering" or "honest responding" condition (Conditions 2 and 4). True population base-rates of malingering are expected to be much lower than 50%, with Slick and colleagues (2004) reporting that two thirds of experts surveyed estimated that the prevalence of *definite* malingering was at least 10%. Only 20% of respondents estimated the prevalence of *possible* malingering to be greater than 30%. As such, the above results present questions about the true diagnostic abilities of the tests, since such tests appear to have had difficulty discriminating between honest responders and those encouraged to malinger, despite having a simulated base-rate of 50% of the population. Possible explanations for this finding will be discussed shortly.

Classification accuracy statistics allow clinicians to make decisions, such as test selection, based on the test's performance capabilities in light of the intended use of the test. A feature of this decision-making process is the trade-off between sensitivity and specificity for a given cut-point of a measure. For example, when a clinician wishes to quickly and efficiently capture individuals who might be at risk for a disease or condition, he or she would ideally select a screening measure. Screening measures are intended to be brief tasks that can be distributed easily to a larger pool of individuals, with the goal of maximizing the number of true positives identified. Following identification from a screener, additional, more rigorous assessment of the condition would be conducted. This diagnostic evaluation would serve as a criterion, identifying true positives and false positives (i.e., those identified by the screener who do not have the condition). Crucially, the proportion of true positives, false positives, true negatives, and false negatives will vary as a function of the chosen cut-point, reflecting the importance of choosing this cut-point wisely.

The present study provided opportunity to evaluate the impact of cut-point selection and the use of a screening measure on classification accuracy. By generating classification accuracy statistics for a variety of cut points for each measure, when discriminating between two experimental conditions, we were able to compare published cut-point recommendations with those suggested to be most optimal in our data. Previous studies have used the following cut-points, as recommended by each test's administration manual: TSI-2 ATR ≥ 15, SIMS > 14, M-FAST ≥ 6, and TOMM < 45. Results from the present study, when using instructional condition as a classifier, indicate that a raw score of 15 on the TSI-2 ATR would yield a sensitivity of approximately 0.17, which would translate to there being a 17% likelihood of a positive test result, given that the individual is truly malingering. Specificity estimates for this cut-score were 0.95. For the SIMS, the sensitivity estimate for an individual with a raw score of 15 was 0.78, while specificity was estimated at 0.28. The M-FAST recommended cut-score of 6 yielded a sensitivity estimate of 0.45 and a specificity estimate of 0.61. The TOMM T2 recommended cut-score of 45 yielded a sensitivity estimate of 0.16, with specificity estimated at 0.91. As the M-FAST and SIMS are intended to be screeners, it is appropriate that liberal (i.e., lower) cut-scores would be used in order to increase the likelihood of capturing true positives. As such, cut points for these measures would ideally yield larger sensitivity estimates.

By plotting sensitivity and specificity for a series of possible cut-points using the present data, we found that the optimal cut-off score for overall accuracy of classification was a TSI-2 ATR score of 7. This cut score correctly classifies 53% of malingerers and 75% of honest responders, for an overall correct classification rate of 65%. This cut score is consistent with the findings of Grey, Elhai, and Briere (2010), who recommend the use of a cut score of 7; however, it is well below the recommended cut-score of 15. For the SIMS, we found that a cut-score of 21 optimized sensitivity (57%) and specificity (51%), for an overall classification accuracy of 54%. This cut-score is much higher than the recommended cut score of 15; however, the use of the SIMS as a screening measure would likely lead to the selection of a cut score that values sensitivity over specificity. Based on present data, the optimal cut score for the M-FAST was 5, which yielded a sensitivity of 54% and a specificity of 56% and an overall classification accuracy of 55%. This cut score is fairly consistent with previous recommendations. Given its difficulty in discriminating between groups and its use of the floor effect, optimal cut scores for the TOMM are more challenging to determine. We found that cut scores ranging from 47 to 49 would yield a sensitivity of 22% (SP = ranging from .91 to .89), with overall classification accuracies hovering near 58%. For the TOMM, as cut scores decline (i.e., failure on more and more items), specificities increase, reflecting how the floor effect increases the probability of a failed test result given honest responding.

Using the SIMS as a dichotomous classifier, instead of condition assignment, allowed for a look into how the use of a screening measure impacts classification accuracy and cut-point selection. In particular, it is noted that using the SIMS to create groups lead to notable changes in the diagnostic classification abilities of the TOMM and M-FAST, while retaining the abilities of the TSI-2 ATR. Under this method, all measures demonstrated statistically significant AUC's. One possible reason for this occurrence is the impact of base rate on predictions. It is known that prediction of low base rate behavior is more difficult than prediction of behaviors that occur commonly, and base rate impacts classification accuracy statistics. As such, by using the SIMS to dichotomize the sample, we succeeded in increasing the "base rate" of malingering beyond that of the conditional assignments. Indeed, increasing sample base rate by screening out those unlikely of meeting criteria is often the goal of screening measures. A second observation is that optimal cut-points often decreased when using the SIMS as a dichotomizer. For example, we found that a cut point of 2 for the TSI-2 ATR and 3 for the M-FAST yielded overall classification accuracies of .69% and 78%, respectively.

Previous investigations of the diagnostic utility of the M-FAST and TSI in a personal injury simulation paradigm found significant, but low, classification accuracy. The present study aimed to expand upon this previous research, utilizing the revised TSI-2 (Alternate version) and additional prominent measures of malingering (the TOMM and SIMS); however, notable discrepancies were found when it comes to the classification accuracies of the TSI and M-FAST In particular, while we found significant classification accuracy of the TSI-2, the M-FAST failed to demonstrate such significant accuracy, yielding an AUC of .53 (p = .36). Discrepancies in the findings of these two studies, despite substantial similarities in methodology, could be due to a the fact that participants in the previous study completed measures and responded to vignettes in person and on paper forms, while participants in the present study did so in person and on computers. This change in method (computer versus paper-and-pencil) may have lead to greater difficulty in recalling and maintaining the case vignette sufficiently enough to feign for the duration of the protocol.

In an attempt to account for inattentive, forgetful, or poorly motivated response patterns, follow-up analyses indicated that a substantial portion of the participant sample had difficulty maintaining specific details about the case vignettes for the duration of the study. Utilizing debriefing questionnaires, we found that less than half of the sample (N = 201; 48%) was able to correctly answer both manipulation check questions. Additionally, it was found that failure on manipulation check questions differed according to condition assignment, suggesting that individuals in Condition 4 were better able to retain information provided. These findings could imply that the information provided in lower levels of manipulation was more difficult to retain, such that individuals in these conditions began responding as if they were in litigation or were suggested to exaggerate symptoms. Additional ROC curves were produced to evaluate whether isolating a sub-sample of participants, who passed manipulation checks, would demonstrate improved classification accuracy. Despite the isolated sample, measures did not demonstrate improved classification ability.

These analyses addressed some concerns associated with whether ambiguity between conditions was driven by the degree to which individuals had difficulty recalling and maintaining information presented in the case vignette, and how such difficulties may have impacted an individuals ability to respond to the measures as directed. A further question to be addressed is the degree to which the manipulation itself was powerful enough to influence a discernable change in response style. As noted by Burris-Garner (2017), who examined the effect of the manipulation in this study, it is possible that case scenarios used evoked endorsments and response styles other than those experimentally intended. For example, it is possible that participants drew more from cognitive symptoms outlined in the vignettes, over psychiatric or affective symptoms. Alternatively, the vignette could have primed individuals to pay more attention to and/or endorse more on items clearly associated with trauma, such as in the TSI-2. Secondly, as evidneced by signficant differences across groups on manipulation checks, it is possible that individuals altered response styles even within manipulation groups with lower degrees of suggestion. This could be due to the fact that information about the purpose of the evaluation was not salient enough, or that common knowledge about motor vehicle accidents and civil litigation overshadowed conditional information. This would contribute to demonstrated ambiguity between the groups. Future studies would identify and focus on a subset of the sample for whom the instructional manipulation worked.

A critical question regarding these results is why prominent measures of malingering (TOMM, M-FAST, and SIMS) generally failed to discriminate between “honest responders” and “malingerers.” When looking at the total score distributions for each test, across Conditions 2 and 4, one can see that the distributions do not differ substantially. Because of this, Chi square tests of difference between the groups indicated no significant differences in response patterns, and ROC curves demonstrated non-significant diagnostic utility (except on the TSI-2 ATR). In summary, the respondents to the tasks typically did not perform as expected and as such, the tests did not operate as expected. Some potential reasons for this outcome have already been discussed; however there a few other factors that may have played a role.

Limitations and Future Research

The present study had several general limitations that warrant discussion. First, this study did not utilize a traditional control-group design. Rather, individuals were assigned to one of four instructional conditions that varied in the degree of suggestion to malinger and in the purpose of the evaluation. The choice to design the study in this way was intentional, as it more accurately reflects a real-world situation, in which group membership is not known prior to assessment and in which respondents typically vary in their clinical presentations across time and testing conditions. As such, while this study design lends greater ecological validity, the absence of a no-condition control limits some conclusions that can be made from the data. In particular, it is not possible to compare those who were told to report various symptoms from those told to respond honestly, as themselves. As such, the present study relied on proxy indicators of "honest responding" (Condition 2; SIMS failure).

The present study also employed a simulation design using an undergraduate sample, which poses two major limitations to the subsequent results. First, the use of a simulation design poses risks in that it is not possible to determine known-groups or to have certainty in group membership. As noted by Rogers and Bender (2013), it is recommended to utilize a known-groups design where possible in order to better capture information about response style and malingering presentations. While logistically very difficult, future research would benefit from the use of ROC analysis within a known-groups, civil litigation population. Secondly, qualitative feedback from the study participants suggests that motivations to faithfully approach each task in the research protocol may have varied. In other simulated malingering research designs, researchers often incentivize subjects in order to increase ecological validity, as individuals involved in actual civil litigation cases are motivated by the opportunity for financial compensation. Similarly, there is the concern for inattentive or careless responding. As noted previously, qualitative feedback during debriefing procedures and failure on manipulation checks suggest that a portion of the sample may have had substantial difficulty retaining the information in the case vignettes and their expectations during the study. Given that participant responses across conditions did not significantly differ for all measures, this suggests the possibility that individuals were unable to retain critical information in the vignette (i.e., suggestion to malinger) that was intended to differentiate conditions.

By using ROC curves as the primary method of data analysis, only two of the four conditions were used to address Hypothesis 2A. This stems from the need for a binary classifier in ROC analysis but means that data from the other two conditions was not addressed. Future directions could include utilizing these conditions to further evaluate the discriminative abilities of each test. For example, in the present study we used the least suggestive and lowest-stakes condition (Condition 2) as a proxy for “honest responding,” with the most suggestive and high-stakes condition (Condition 4) as a proxy for “malingering.” Data from conditions 1 and 3, which have iteratively more suggestion to malinger and greater implied stakes associated with the evaluation could be used to demonstrate that a measure can distinguish not only between honest responders and malingerers, but also between malingerers with or without the suggestion to malinger or the relative impact of litigation on an individual’s inclination to feign.

As noted by Boone (2011), it is not atypical for honestly-responding individuals to fail a single malingering measure. It is, however, unusual for individuals to fail more than one measure, and iteratively rarer for someone to fail greater than two. This finding supports recommendations to improve assessment validity by utilizing more than one malingering measure, when dishonest responding is suspected. While survey research indicates that most clinicians do use more than one measure, there is considerable variation in the overall number and type of tests used, both across clinicians and within individual practices (Boccaccini and Brodsky, 1999). Such variability, in conjunction with findings that multiple test failures are rare, suggests that that classification accuracy of individual malingering tests could be further evaluated by determining the additive impact of successive test failures, and identifying a recommendation for the number of tasks or number of failures that would indicate a high likelihood of malingering. Future research utilizing this data set could examine participant profiles across the given measures, looking at which individuals demonstrated failure on only one task versus those who failed two, three, or all four. Determining which threshold allows for the most accurate classification would aid clinicians in test selection and assessment planning.

The present study evaluated the interrelationships amongst SVT and PVT's of malingering using a multi-trait, multi-method matrix. This matrix of correlations allows for researchers to examine construct validity by understanding the relationships between tests of a given construct and tests with a particular administration type. In this study, we found evidence consistent with the conceptualization of malingering as a non-unitary construct. Given the number of available malingering measures (for both symptom and performance validity), future research should continue to address the inter-relationships amongst these tasks in various medico-legal contexts. Future analyses could also look at the magnitude and direction of relationships between malingering measures at the subscale level. For example, it would be valuable to know whether reported memory impairment on SVT measures correlates as expected with actual memory impairment on PVT measures. Furthermore, more sophisticated methods of data analysis (e.g., structural equations modeling; SEM) would allow for a deeper look into these relationships. Methods such as SEM allow for the partitioning out of variance due to method, shared constructs, and other variables, which could further elucidate the broader construct of malingering.

References

Ackerman, M., & Kane, A. (1998). *Psychological Experts in Personal Injury Actions.* Aspen Law & Business.

Alwes, Y., Clark, J., Berry, D., & Granacher, R. (2008). Screening for feigning in a civil forensic setting. *Journal of Clinical and Experimental Neuropsychology*, 133-140.

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U. S.). (2014). *Standards for Educational and Psychological Testing.*

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (Fourth, Text Revision ed.). Washington, D. C. : Author.

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

Arce, R., Farina, F., & Buela, G. (2008). Assessing and Detecting the Ability to Faking Psychological Injury as a Consequence of a Motor Vehicle Accident on the MMPI-2 Using Mock Victims. *Revista Latinoamericana de Psicologia*, 485-496.

Berry, D. R., & Nelson, N. (2010). DSM-5 and Malingering: A Modest Proposal. *Psychological Injury and Law*.

Boccaccini, M., & Brodsky, S. (1999). Diagnostic Test Usage by Forensic Psychologists in Emotional Injury Cases. *Professional Psychology: Research and Practice*, 253-259.

Boone, K. (2011). Clarification or confusion? A Review of Rogers, Bender, and Johnson's a critical analysis of the MND criteria. *Psychological Injury and Law*, 157-162.

Breslau, N. (2001). The Epidemiology of Posttraumatic Stress Disorder: What is the Extent of the Problem? *Journal of Clinical Psychiatry*, 16-22.

Briere, J. (1995). *Trauma Symptom Inventory Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Briere, J. (2011). *Trauma Symptom Inventory - 2: Professional Manual.* Lutz: PAR.

Burris Garner, E. (2017). Comparison of Malingering Measures, Suggestion, and Litigation Status Among Personal Injury Litigants. *Dissertation*.

Butcher, J., Dahlstrom, W., Graham, J., Tellegen, A., & Kreammer , B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Manual for Administration and Scoring.* Minneapolis, MN: University of Minneapolis Press.

Campbell, D., & Fiske, D. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 81-105.

Christiansen, A., & Vincent, J. (2012). Assessment of Litigation Context, Suggestion, and Malingering Measures among Simulated Personal Injury Litigants . *Journal of Forensic Psychology Practice*, 238-258.

Cohen, T. H. (2004). *Tort Trial and Verdicts in Large Counties, 2001.* U.S. Department of Justice, Office of Justice Programs.

Demakis, G., Gervais, R., & Rohling, M. (2008). The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *Clinical Neuropsychology*, 879-895.

Dobie, D. J., Kivlahan, D. R., Maynard, C., Bush, K. R., McFall, M., Epler, A. J., & Bradley, K. A. (2002). Screening for post-traumatic stress disorder in female Veteran's Affairs patients: validation of the PTSD checklist. *General Hospital Psychiatry*, 367-374.

Elhai, J., Gray, M., Naifeh, J., Butcher, J., Davis, J., Falsetti, S., & Best, C. (2005). Utility of the Trauma Symptom Inventory's Atypical Response Scale in Detecting Malingered Post-Traumatic Stress Disorder. *Assessment*, 210-219.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 861-874.

Foote, W., & Lareau, C. (2013). Psychological Evaluation of Emotional Damages in Tort Cases. In R. Otto, & I. Weiner, *Handbook of Psychology: Forensic Psychology, Second Edition* (pp. 172-200). Hoboken, NJ: John Wiley & Sons, Inc. .

Goodwin, L., & Leech, N. (2003). The Meaning of Validity in the New Standards for Educational and Psychological Testing: Implications for Measurement Courses. *Measurement and Evaluation in Counseling and Development*, 181-191.

Gray, M. J., Elhai, J. D., & Briere, J. (2010). Evaluation of the Atypical Response scale of the Trauma Symptom Inventory-2 in detecting simulated post-traumatic stress disorder. *Journal of Anxiety Disorders*, 447-451.

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

Greiner, M., Pfeiffer, D., & Smith, R. D. (2000). Principles and practical application of the reciever-operating characteristic analysis for diagnostic tests. *Preventitive Veternary Medicine*, 23-41.

Guriel-Tennant, J., & Fremouw, W. (2006). Impact of trauma history and coaching on malingering of posttraumatic stress disorder using the PAI, TSI, and M-FAST. *The Journal of Forensic Psychiatry and Psychology*, 577-592.

Guy, L., Kwartner, M., & Miller, H. (2006). Investigating the M-FAST: Psychometric Properties and Utility to Detect Diagnostic Specific Malingering. *Behavioral Sciences & the Law*, 687-702.

Guy, L., Kwartner, P., & Miller, H. (2006). Investigating the M-FAST: Psychometric Properties and Utility to Detect Diagnosis Specific Malingering. *Behavioral Sciences and the Law*, 687-702.

Hall, R. W., & Hall, R. W. (2007). Detection of Malingered PTSD: An Overview of Clinical, Psychometric, and Physiological Assessment: Where Do We Stand? *Journal of Forensic Sciences*, 717-725.

Klotz Flitter, J. M., Elhai, J. D., & Gold, S. N. (2003). MMPI-2 F scale elevations in adult victims of child sexual abuse. *Journal of Traumatic Stress*, 269-274.

Langton, L., & Cohen, T. (2008). *Civil Bench and Jury Trials in State Courts, 2005.* U.S. Department of Justice, Office of Justice Programs.

Larrabee, G. (2003). Detection of Malingering using neuropsychologically atypical performance patterns on standard neuropsychological tests. *Clinical Neuropsychologist*, 410-425.

Larrabee, G., & Berry, D. T. (2007). Diagnostic Classification Statistics and Diagnostic Validity of Malingering Assessment. In G. Larrabee, *Assessment of Malingered Neuropsychological Deficits* (pp. 14-26). New York : Oxford University Press.

Less-Haley, P. R., & Dunn, J. T. (1994). The Ability of Naive Subjects to Report Symptoms of Mild Brain Injury, Post-Traumatic Stress Disorder, Major Depression, and Generalized Anxiety Disorder. *Journal of Clinical Psychology*, 252-256.

Lezak, M. (1995). *Neuropsychological Assessment, Third Edition.* New York, NY: Oxford University Press.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological Evaluations for the Courts (Third Edition).* New York: The Guilford Press.

Merckelbach, H., & Smith, G. (2003). Diagnostic Accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, 145-152.

Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning. *American Psychologist*, 741-749.

Miller, H. A. (2001). *M-FAST: Miller Forensic Assessment of Symptoms Test Professional Manual.* Odessa, FL: Psychological Assessment Resources, Inc. .

Morey, L. (2007). *Personality Assessment Inventory (PAI): Professional Manual.* PAR (Psychological Assessment Resources).

Neal, T., & Grisso, T. (2014). Assessment Practices and Expert Judgment Methods in Forensic Psychology and Psychiatry. *Criminal Justice and Behavior*, 1406-1421.

Parks, A. (2015). Using Performance Validity and Symptom Validity Tests to Detect Feigned Postconcussion and Posttraumatic Stress Symptomatologies: A Simulation Study. *Dissertation*.

Peace, K. A., & Masliuk, K. A. (2011). Do Motivations for Malingering Matter? Symptoms of Malingered PTSD as a Function of Motivation and Trauma Type. *Psychological Injury and Law*, 44-55.

Pintea, S., & Moldovan, R. (2009). The Reciever-Operating Characteristic (ROC) Analysis: Fundamentals and Applications in Clinical Psychology. *Journal of Cognitive and Behavioral Psychotherapies*, 49-66.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedis. *Journal of Applied Psychology*, 879-903.

Purtle, J., Lynn, K., & Malik, M. (2016). "Calculating the Toll of Trauma" in the Headlines: Portrayals of Posttraumatic Stress Disorder in the New York Times (1980-2015). *American Journal of Orthopsychiatry*, 632-638.

Resnick, P. (1997). Malingering of Posttraumatic Disorders. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (Second ed., pp. 130-152). New York: The Guilford Press.

Resnick, P., West, S., & Payne, J. (2008). Malingering of Posttraumatic Disorders. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (pp. 109-127). New York: The Guilford Press.

Rogers, R. (2008). An Introduction to Response Styles. In R. Rogers, *Clinical Assessment of Malingering and Deception, Third Edition.* New York, NY: The Guilford Press.

Rogers, R. (2008). Detection Strategies for Malingering and Defensiveness. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (pp. 14-35). New York: The Guilford Press.

Rogers, R., & Bender, S. (2013). Evaluation of Malingering and Related Response Styles. In R. Otto, & I. Weiner (Eds.), *Handbook of Psychology, Volume 11: Forensic Psychology* (pp. 517-540). Hoboken: John Wiley & Sons, Inc.

Rogers, R., Bagby, R., & Dickens, S. (1992). *Structured interview of reported symptoms (SIRS) and professional manual.* Odessa, FL: Psychological Assessment Resources.

Schultz, I. (2003). The Relationship between Psychological Impairmetn and Occupational Disability. In I. Schultz, & D. Brady, *Psychological Injuries at Trial* (pp. 65-101). Chicago, IL: American Bar Association.

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

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

Smith, G., & Burger, G. (1997). Detection of Malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law*, 183-189.

Stone, A. A. (1993). Post-Traumatic Stress Disorder and the Law: Critical Review of the New Frontier. *The Bulletin of the American Academy of Psychiatry and the Law*, 23-36.

Story, D. L. (2000). Validation of a Short Form of the Structured Interview of Reported Symptoms (SIRS).

Taylor, S., Frueh, B., & Asmundson, G. (2007). Detection and management of malingering in people presenting for treatment of posttraumatic stress disorder: Methods, obstacles, and recommendations. *Journal of Anxiety Disorders*, 22-41.

Thomas, T., & Fremouw, W. (2009). The Stroop interference effect and spontaneous recall in true and malingered motor-vehicle accident-related PTSD. *Journal of Forensic Psychiatry & Psychology*, 936-949.

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

Tombaugh, T. N. (2006). *Test of Memory Malingering (TOMM); Research Monograph: Summary Application in Clinical and Research Settings.* North Tonawanda: Multi-Health Systems, Inc. .

Vallano, J. (2013). Psychological Injuries and Legal Decision Making in Civil Cases: What We Know and What We Do not Know. *Psychological Injury and Law*, 99-112.

van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Invetory of Malingered Symptomatology (SIMS): A Systematic Review and Meta-Analysis. *The Clinical Neuropsychologist*, 1336-1365.

Vincent, J. P., & Christiansen, A. K. (2012). Assessment of Litigation Context, Suggestion, and Malingering Measures among Simulated Personal Injury Litigants. *Journal of Forensic Psychology Practice*, 238-258.

Vitacco, M., Rogers, R., Gabel, J., & Munizza, J. (2007). An evaluation of malingering screens with competency to stand trial patients: a known-groups comparison. *Law and Human Behavior*, 249-260.

Widows, M., & Smith, G. (2005). *SIMS: Structured Inventory of Malingered Symptomatology.* Lutz, FL: PAR, Inc. .

Williams, C., Lees-Haley, P., & Djanogly, S. (1999). Clinical Scrutiny of Litigants' Self-Reports. *Professional Psychology: Research and Practice*, 361-367.

Wisdom, N., Callahan, J., & Shaw, T. (2010). Diagnostic Utility of the Structured Inventory of Malingered Symtomatology to Detect Malingering in a Forensic Sample. *Archives of Clinical Neuropsychology*, 118-125.

Young, G., Kane, A., & Nicholson, K. (2007). *Causality of Psychological Injury: Presenting Evidence in Court.* New York, NY: Springer Science + Business Media, LLC.

Zoellner, L., Pruitt, L., Farach, F., & Jun, J. (2014). Understanding Heterogeneity in PTSD: Fear, Dysphoria, and Distress. *Depression & Anxiety*, 97-106.

Tables and Figures

Table 1. *Descriptive Statistics by Full Sample and Condition*

	Full Sample	Condition						
		1	2	3	4			
	M (SD) or *n* (%)	*M* (SD) or *n* (%)	*M* (SD) or *n* (%)	*M* (SD) or *n* (%)	*M* (SD) or *n* (%)	*F* or X^2	*(df)*	*p*
Total Participants	411	105	104	106	96			
Age	23.0 (5.59)	22.4 (4.59)	23.1 (5.77)	23.9 (6.91)	22.65 (4.62)	1.51	(3, 407)	0.212
Gender						2.97	(3)	0.397
Female	283 (68.9)	75 (71.4)	73 (70.2)	66 (62.3)	69 (71.9)			
Male	128 (31.1)	30 (28.6)	31 (29.8)	40 (37.7)	27 (28.1)			
Race/Ethnicity						12.69	(12)	0.392
African-American	66 (16.1)	20 (19.0)	16 (15.4)	14 (13.2)	16 (16.7)			
Asian-American	95 (23.1)	21 (20.0)	31 (29.8)	27 (25.5)	16 (16.7)			
Caucasian	94 (22.9)	19 (18.1)	22 (21.2)	27 (25.5)	26 (27.1)			
Hispanic	115 (28.0)	36 (34.3)	28 (26.9)	26 (24.5)	25 (26.0)			
Other/Multiracial	41 (10.0)	9 (8.6)	7 (6.7)	12 (11.3)	13 (13.5)			
Academic Standing						8.59	(12)	0.737
Freshman	43 (10.5)	12 (11.4)	12 (11.5)	9 (9.5)	10 (10.4)			
Sophomore	81 (19.7)	20 (19.0)	22 (21.2)	24 (22.6)	15 (15.6)			
Junior	173 (42.1)	44 (41.9)	41 (39.4)	48 (45.3)	40 (41.7)			
Senior	102 (24.8)	23 (21.9)	26 (25.0)	24 (22.6)	29 (30.2)			
Post-baccalaureate	12 (2.9)	6 (5.7)	3 (2.9)	1 (0.9)	2 (2.1)			
Past MVA Involvement†	147 (33.3)	33 (31.4)	33 (31.7)	37 (34.9)	34 (35.4)	0.6	(3)	0.897

** indicates p-value significant at the .01 level. * indicates p-value significant at the .05 level. † reflects participants who indicated past involvement in a motor vehicle accident, which was not the participants fault, but in which they sustained psychological or physical injuries.

Table 2. *Testing Statistics by Full Sample and Condition*

		Condition			
Measure	Full Sample	1	2	3	4
TSI-2 ATR					
Mean (SD)	5.75 (6.19)	4.89 (5.60)	4.31 (5.11)	6.00 (6.45)	7.98 (6.97)
Minimum	0	0	0	0	0
Maximum	24	22	19	24	24
Failure Rate (Cut-off: ≥15)	10.7%	6.7%	4.8%	15.1%	16.7%
SIMS					
Mean (SD)	23.47(12.68)	21.10 (11.55)	22.29 (10.88)	24.53 (14.07)	26.17 (13.59)
Minimum	2	2	2	2	4
Maximum	69	53	48	69	61
Failure Rate (Cut-off: >14)	71.0%	61.9%	72.1%	72.6%	78.1%
M-FAST					
Mean (SD)	5.31 (4.78)	4.61 (4.64)	5.12 (4.35)	5.71 (5.20)	5.86 (4.87)
Minimum	0	0	0	0	0
Maximum	24	21	19	24	19
Failure Rate (Cut-off: ≥6)	39.7%	32.4%	38.5%	43.4%	44.8%
TOMM T2					
Mean (SD)	48.25 (5.82)	48.79 (5.19)	48.18 (6.50)	48.84 (4.29)	47.09 (6.97)
Minimum	11	11	13	27	16
Maximum	50	50	50	50	50
Failure Rate (Cut-off: <45)	8.5%	4.8%	8.7%	5.7%	15.6%

Note: Failure rates reflect the proportion of participants in the sample scoring higher or lower than the measures recommended cut-off.

Table 3. *Independent Samples T-tests between Total Score Means of Conditions 2 and 4*

	Condition							
	2 (*n* = 105)		4 (*n* = 96)					
	M	SD	M	SD	t	df	p-value	95% CI [L, U]
TSI-2 ATR	4.31	5.11	7.98	6.97	-4.27	198	.000**	-5.37, -1.96
SIMS	22.29	10.88	26.17	13.59	-2.24	198	.027*	-7.3, -0.46
M-FAST	5.12	4.35	5.86	4.87	-1.15	198	.252	-2.03, 0.54
TOMM T2	48.18	6.5	47.09	6.97	1.14	198	.254	-0.79, 2.97

** indicates p-value significant at the .01 level. * indicates p-value significant at the .05 level.

Table 4. *Chi Square of Conditions 2 and 4*

	Condition 2		Condition 4			
	n (%)		*n* (%)			
	Pass	Fail	Pass	Fail	X^2	*p*
TSI-2 ATR (Cut-off: ≥15)	99 (95.2)	5 (4.8)	80 (83.3)	16 (16.7)	7.47	0.006**
SIMS (Cut-off: >14)	29 (27.9)	75 (72.1)	21 (21.9)	75 (78.1)	0.96	0.327
M-FAST (Cut-off: ≥6)	64 (61.5)	40 (38.5)	53 (55.2)	43 (44.8)	0.82	0.364
TOMM T2 (Cut-off: <45)	95 (91.3)	9 (8.7)	81 (84.4)	15 (15.6)	2.29	0.130

** indicates p-value significant at the .01 level. * indicates p-value significant at the .05 level.

Condition 2 sample size = 104; Condition 4 = 96.

Table 5. *Bivariate Correlations between SVT and PVT Total Scores*

		Measure			
		TSI-2 ATR	SIMS	M-FAST	TOMM T2
SVT's	TSI-2 ATR	-			
	SIMS	.65**	-		
	M-FAST	.51**	.62**	-	
PVT	TOMM T2†	-0.28**	-0.26**	-0.15**	-

** indicates correlation is significant at the $p < .01$ level (two-tailed).

† Lower scores on the TOMM are indicative of malingering, as opposed to SVT measures, where higher scores are more indicative of malingered responding.

Figure 1. *Receiver operating characteristics (ROC) curve for the TSI-2-A ATR, dichotomized by condition.*

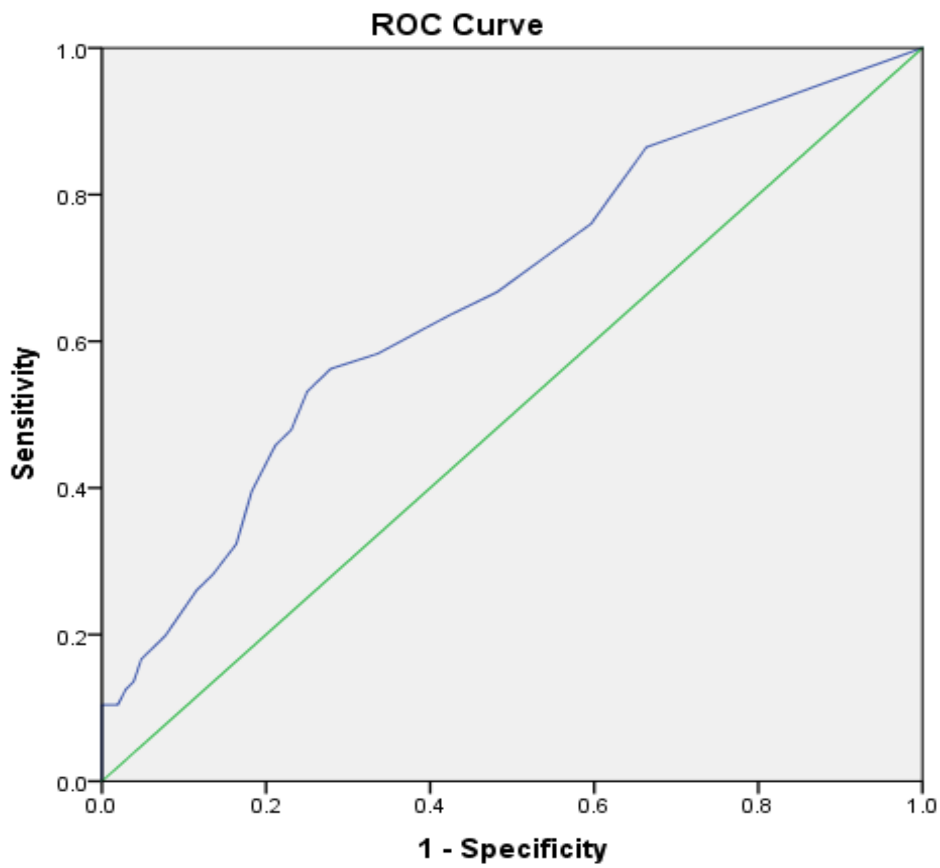

Figure 2. *Receiver operating characteristics (ROC) curve for the SIMS, dichotomized by condition.*

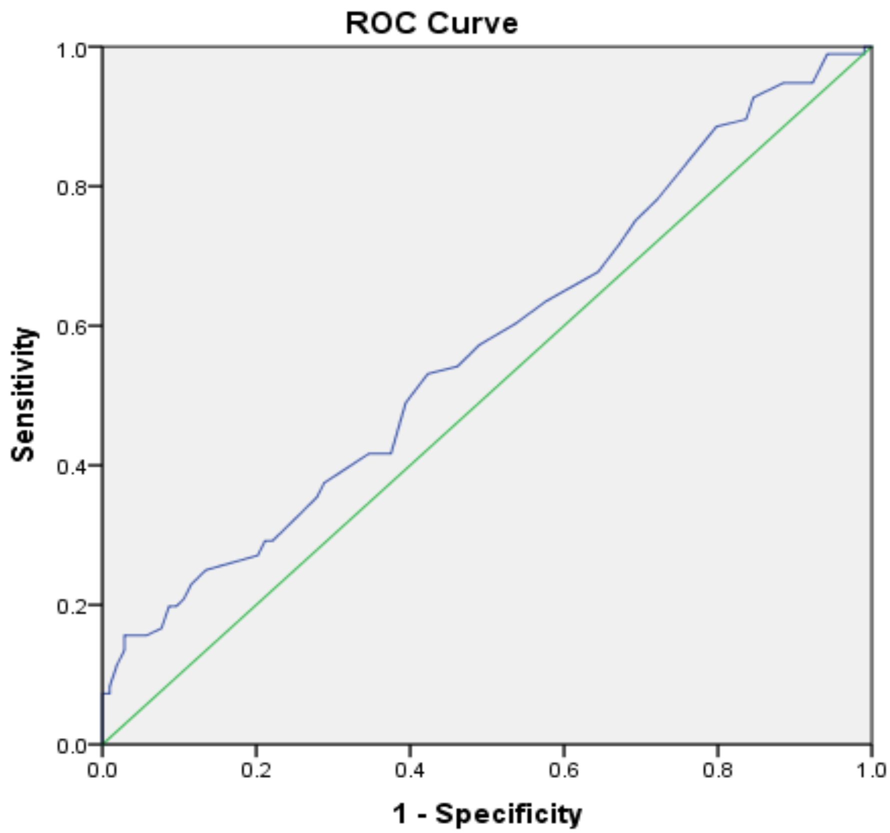

Figure 3. *Receiver operating characteristics (ROC) curve for the M-FAST, dichotomized by condition.*

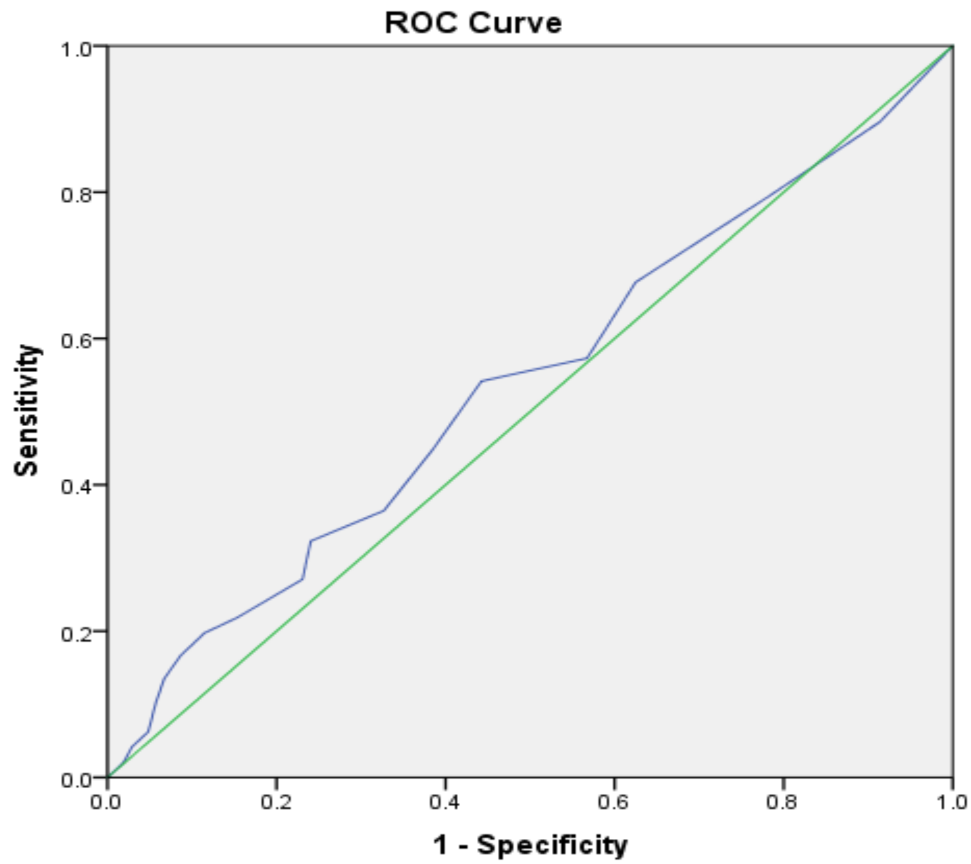

Figure 4. *Receiver operating characteristics (ROC) curve for the TOMM Trial 2, dichotomized by condition.*

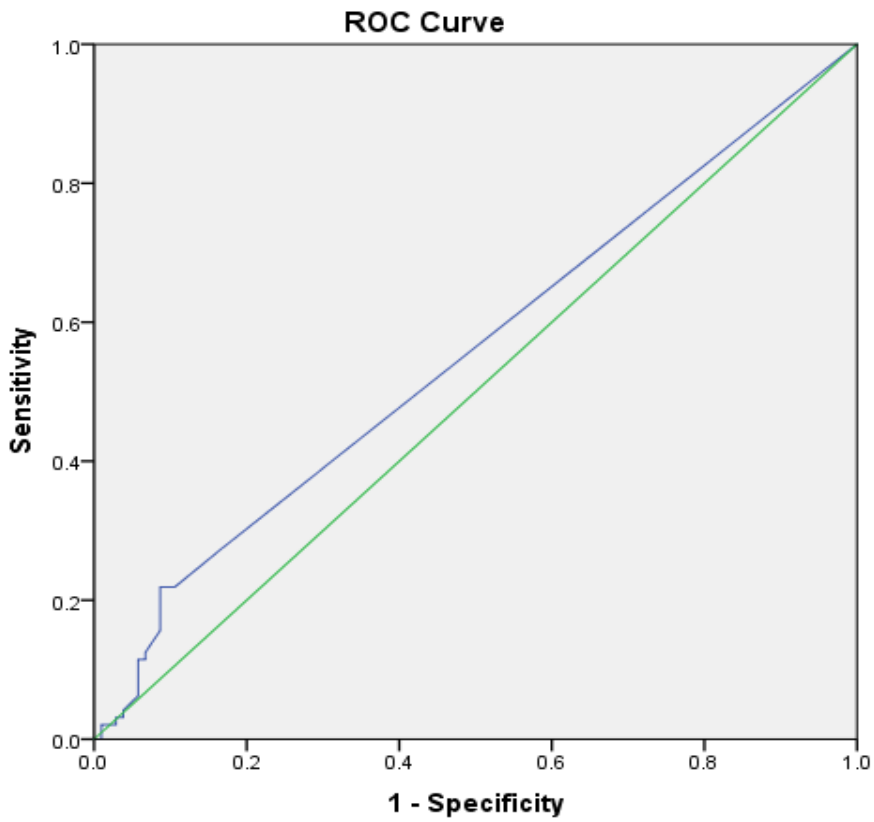

Table 6. *AUC Summaries by Test based on ROC Curves dichotomized by condition*

	AUC	SE	Sig.	95% CI [L, U]
TSI-2 ATR	.662	.038	.000**	.587, .737
SIMS	.571	.040	.083	.492, .650
M-FAST	.537	.041	.362	.457, .618
TOMM T2	.555	.041	.176	.475, .635

** indicates p-value significant at the .01 level. AUC = area under the ROC curve; SE = standard error; Sig. = asymptotic significance, where H0 = 0.5; CI = confidence interval, with Lower and Upper bounds.

Figure 5. *Sensitivity and Specificity plotted against different cut-scores on the TSI-2-A ATR*

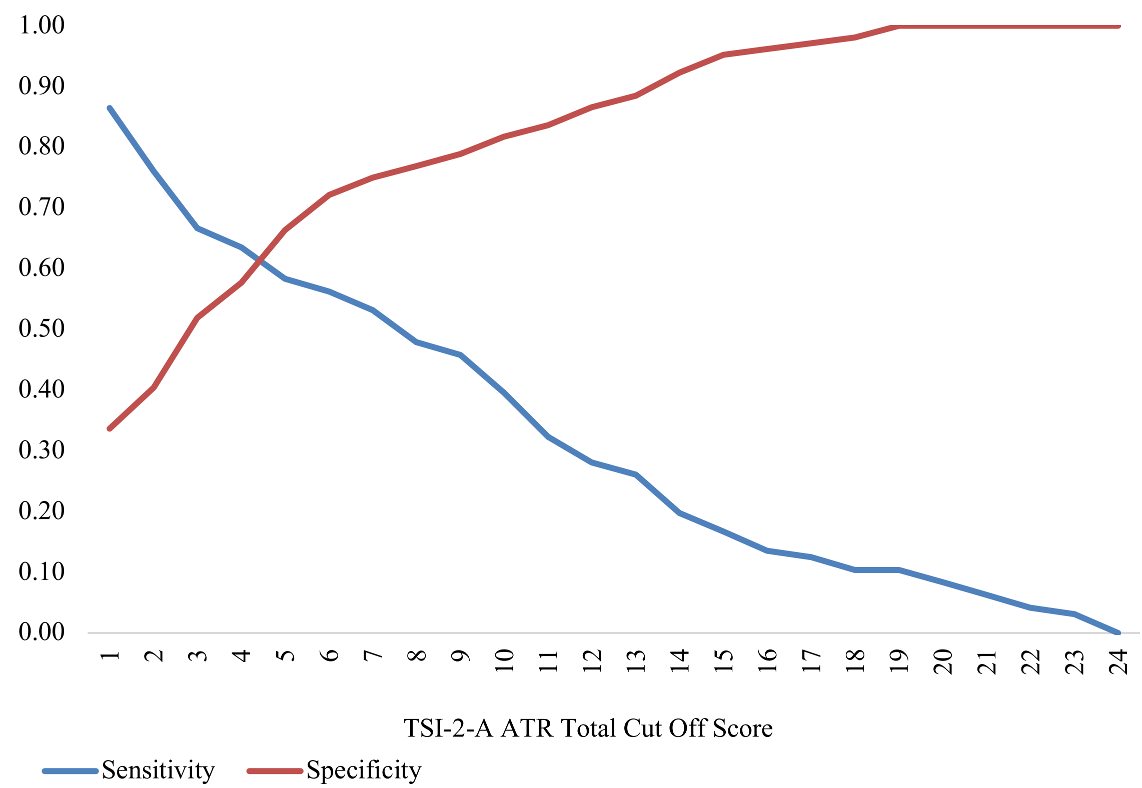

Figure 6. *Sensitivity and Specificity plotted against different cut-scores on the SIMS*

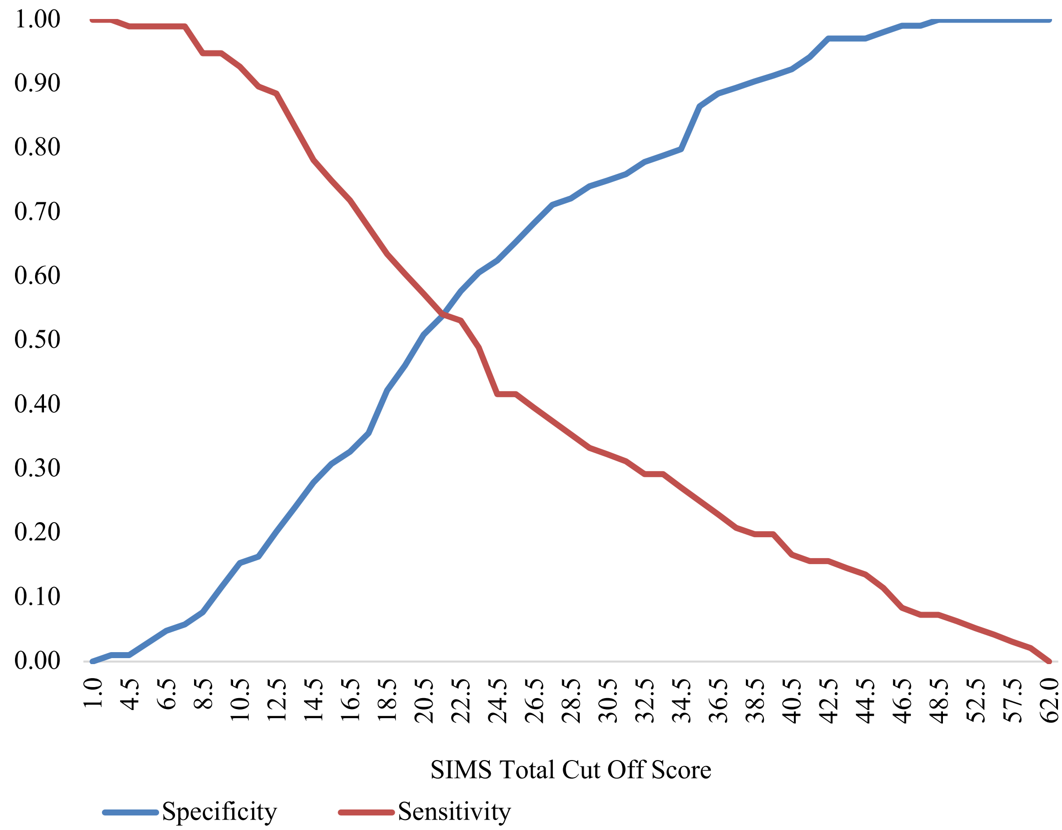

Figure 7. *Sensitivity and Specificity plotted against different cut-scores on the M-FAST*

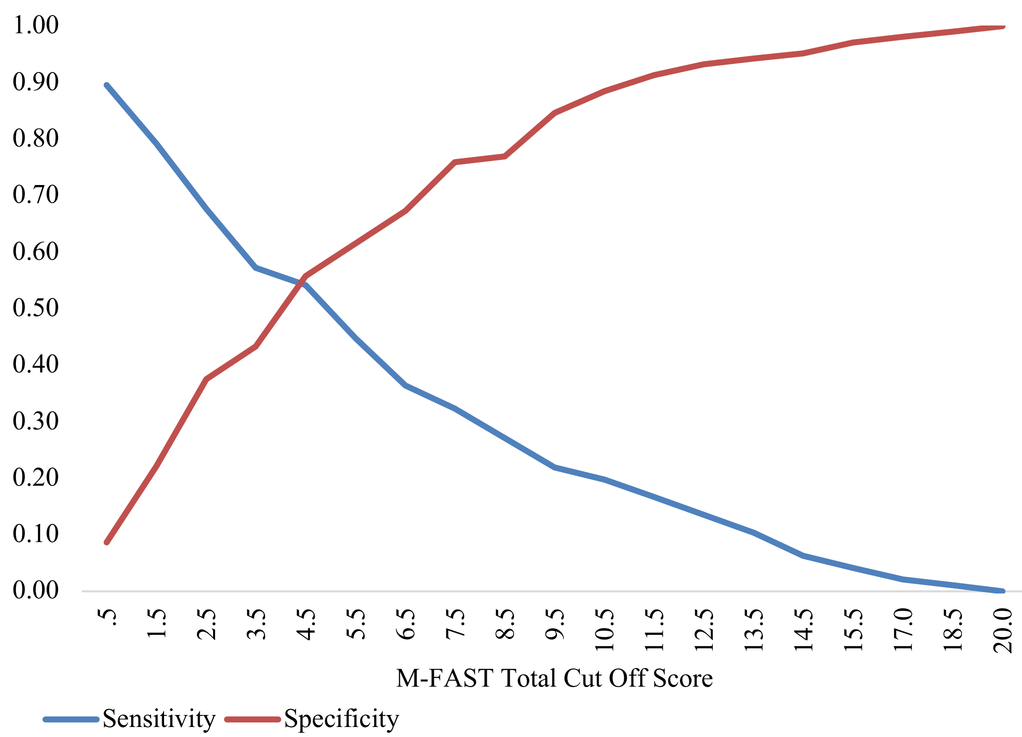

Figure 8. *Sensitivity and Specificity plotted against different cut-scores on the TOMM T2*

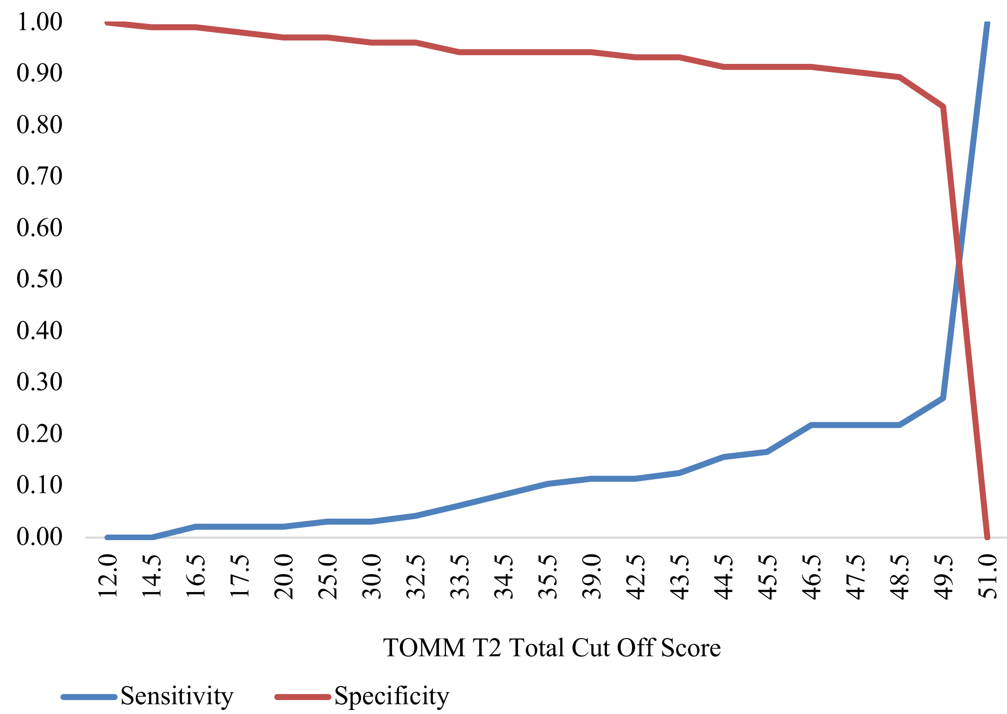

Table 7. *Classification Accuracy Statistics across Observed Scores for the TSI-2-A ATR*

Cut-Score	Sensitivity	Specificity	FPR	PPP	NPP	ACC
3	0.67	0.52	0.48	0.56	0.63	0.59
5	0.58	0.66	0.34	0.62	0.63	0.63
7	0.53	0.75	0.25	0.66	0.63	0.65
9	0.46	0.79	0.21	0.67	0.61	0.63
11	0.32	0.84	0.16	0.65	0.57	0.59
13	0.26	0.88	0.12	0.68	0.56	0.59
15	0.17	0.95	0.05	0.76	0.55	0.58
17	0.13	0.97	0.03	0.80	0.55	0.57
19	0.10	0.98	0.02	0.83	0.54	0.56

Note. Cut-score = observed raw score. Sensitivity = probability of a positive test result given the condition is present (malingering). Specificity = probability of a negative test result, given the condition is absent (not malingering). FPR = False Positive Rate, probability that an individual recieves a positive test result, depite the condition being absent . PPP = Positive Predictive Power, probability the condition is present when the test is positive. NPP = Negative Predictive Power, probability the condition is not present when the test is negative. ACC = Accuracy, proportion of correctly classified individuals to the total sample. All calculations are based on the conditional base rate of 48%.

Table 8. *Classification Accuracy Statistics across Observed Scores for the SIMS*

Cut-Score	Sensitivity	Specificity	FPR	PPP	NPP	ACC
10	0.95	0.12	0.88	0.50	0.71	0.52
11	0.93	0.15	0.85	0.50	0.70	0.53
12	0.90	0.16	0.84	0.50	0.63	0.52
13	0.89	0.20	0.80	0.51	0.66	0.53
14	0.83	0.24	0.76	0.50	0.61	0.53
15	0.78	0.28	0.72	0.50	0.58	0.52
16	0.75	0.31	0.69	0.50	0.57	0.52
17	0.72	0.33	0.67	0.50	0.56	0.52
18	0.68	0.36	0.64	0.49	0.54	0.51
19	0.64	0.42	0.58	0.50	0.56	0.53
20	0.60	0.46	0.54	0.51	0.56	0.53
21	0.57	0.51	0.49	0.52	0.56	0.54
22	0.54	0.54	0.46	0.52	0.56	0.54
23	0.53	0.58	0.42	0.54	0.57	0.56

Note. Cut-score = observed raw score. Sensitivity = probability of a positive test result given the condition is present (malingering). Specificity = probability of a negative test result, given the condition is absent (not malingering). FPR = False Positive Rate, probability that an individual recieves a positive test result, depite the condition being absent . PPP = Positive Predictive Power, probability the condition is present when the test is positive. NPP = Negative Predictive Power, probability the condition is not present when the test is negative. ACC = Accuracy, proportion of correctly classified individuals to the total sample. All calculations are based on the conditional base rate of 48%.

Table 9. *Classification Accuracy Statistics across Observed Scores for the M-FAST*

Cut-Score	Sensitivity	Specificity	FPR	PPP	NPP	ACC
2	0.79	0.22	0.78	0.48	0.53	0.50
3	0.68	0.38	0.63	0.50	0.56	0.52
4	0.57	0.43	0.57	0.48	0.52	0.50
5	0.54	0.56	0.44	0.53	0.57	0.55
6	0.45	0.62	0.38	0.52	0.55	0.54
7	0.36	0.67	0.33	0.51	0.53	0.53
8	0.32	0.76	0.24	0.55	0.55	0.55
9	0.27	0.77	0.23	0.52	0.53	0.53
10	0.22	0.85	0.15	0.57	0.54	0.55

Note. Cut-score = observed raw score. Sensitivity = probability of a positive test result given the condition is present (malingering). Specificity = probability of a negative test result, given the condition is absent (not malingering). FPR = False Positive Rate, probability that an individual recieves a positive test result, depite the condition being absent . PPP = Positive Predictive Power, probability the condition is present when the test is positive. NPP = Negative Predictive Power, probability the condition is not present when the test is negative. ACC = Accuracy, proportion of correctly classified individuals to the total sample. All calculations are based on the conditional base rate of 48%.

Table 10. *Classification Accuracy Statistics across Observed Scores for the TOMM T2*

Cut-Score	Sensitivity	Specificity	FPR	PPP	NPP	ACC
40	0.11	0.94	0.06	0.65	0.54	0.55
41	0.11	0.94	0.06	0.65	0.54	0.55
42	0.11	0.94	0.06	0.65	0.54	0.55
43	0.11	0.93	0.07	0.61	0.53	0.54
44	0.13	0.93	0.07	0.63	0.54	0.55
45	0.16	0.91	0.09	0.63	0.54	0.55
46	0.17	0.91	0.09	0.64	0.54	0.56
47	0.22	0.91	0.09	0.70	0.56	0.58
48	0.22	0.90	0.10	0.68	0.56	0.58
49	0.22	0.89	0.11	0.66	0.55	0.57
50	0.27	0.84	0.16	0.60	0.55	0.57

Note. Cut-score = observed raw score. Sensitivity = probability of a positive test result given the condition is present (malingering). Specificity = probability of a negative test result, given the condition is absent (not malingering). FPR = False Positive Rate, probability that an individual recieves a positive test result, depite the condition being absent . PPP = Positive Predictive Power, probability the condition is present when the test is positive. NPP = Negative Predictive Power, probability the condition is not present when the test is negative. ACC = Accuracy, proportion of correctly classified individuals to the total sample. All calculations are based on the conditional base rate of 48%.

Figure 9. *ROC curve for the TSI-2-A ATR, dichotomized by SIMS Pass/Failure (Cut-off > 14).*

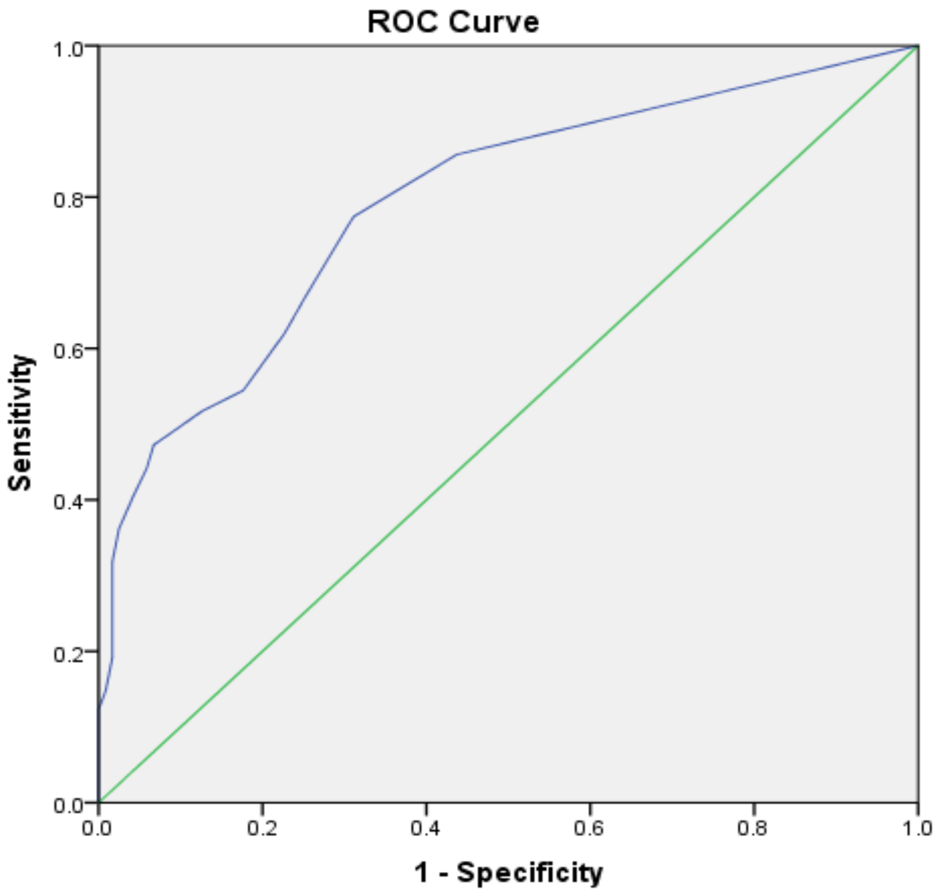

Figure 10. *ROC curve for the M-FAST, dichotomized by SIMS Pass/Failure (Cut-off > 14).*

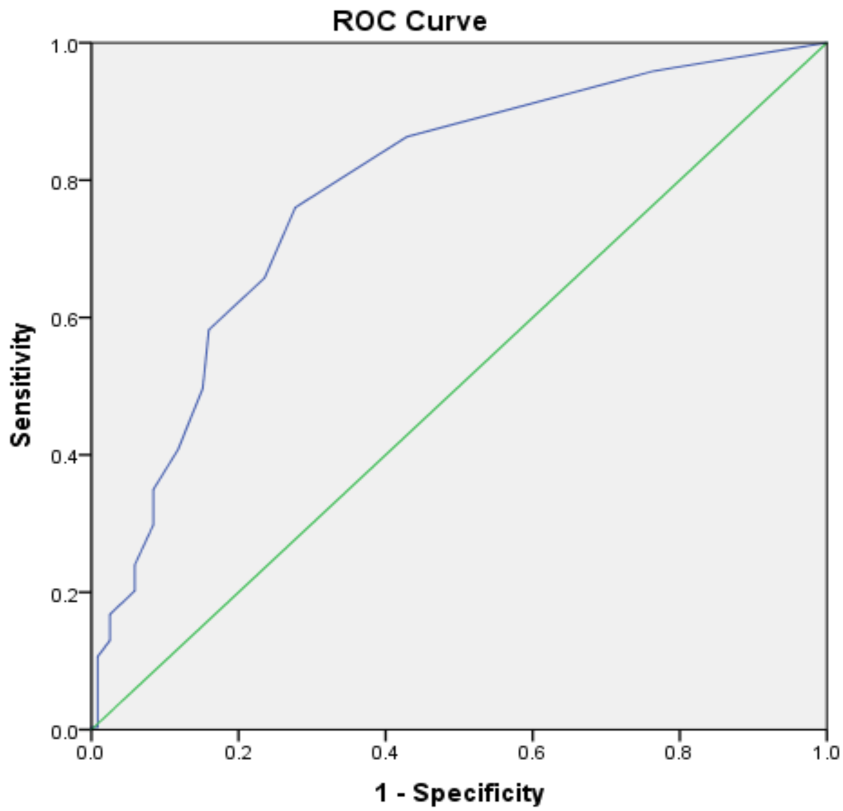

Figure 11. *ROC curve for the TOMM Trial 2, dichotomized by SIMS Pass/Failure (Cut-off > 14).*

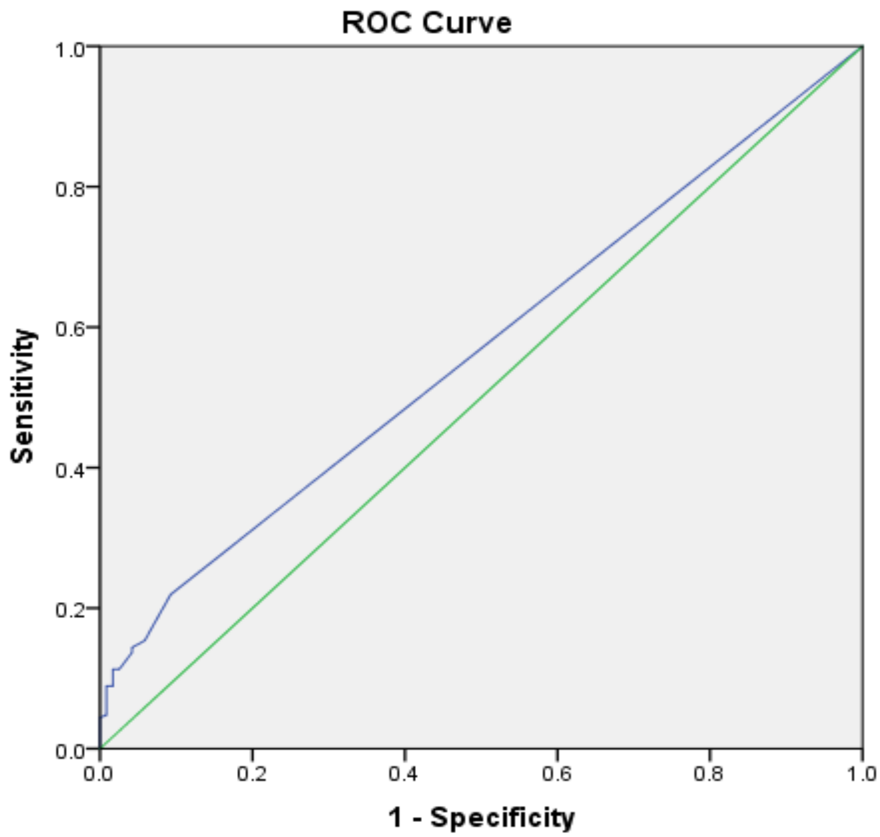

Table 11. *AUC Summaries by Test based on ROC Curves dichotomized by SIMS pass/failure*

	AUC	SE	Sig.	95% CI [L, U]
TSI-2 ATR	.792	.023	.000**	.747, .837
M-FAST	.784	.026	.000**	.733, .834
TOMM T2	.566	.030	.035*	.508, .625

** indicates p-value significant at the .01 level. * indicates p-value significant at the .05 level. AUC = area under the ROC curve; SE = standard error; Sig. = asymptotic significance, where H0 = 0.5; CI = confidence interval, with Lower and Upper bounds.

Figure 12. *Sensitivity and Specificity plotted against different cut-scores on the TSI-2-A ATR*

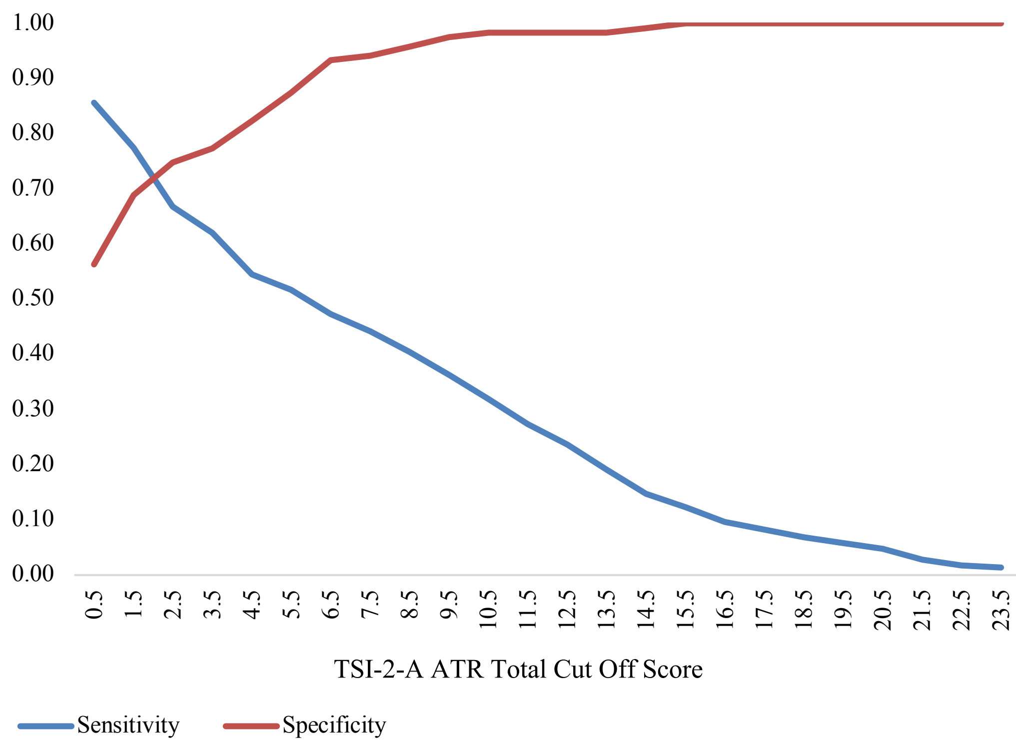

Figure 13. *Sensitivity and Specificity plotted against different cut-scores on the M-FAST*

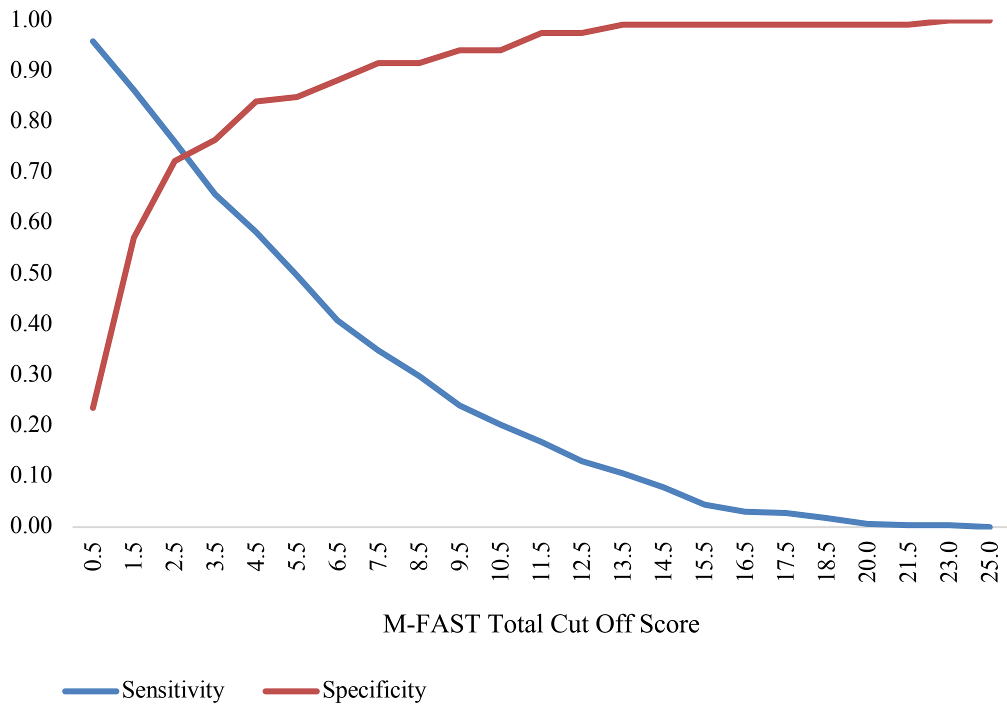

Figure 14. *Sensitivity and Specificity plotted against different cut-scores on the TOMM T2*

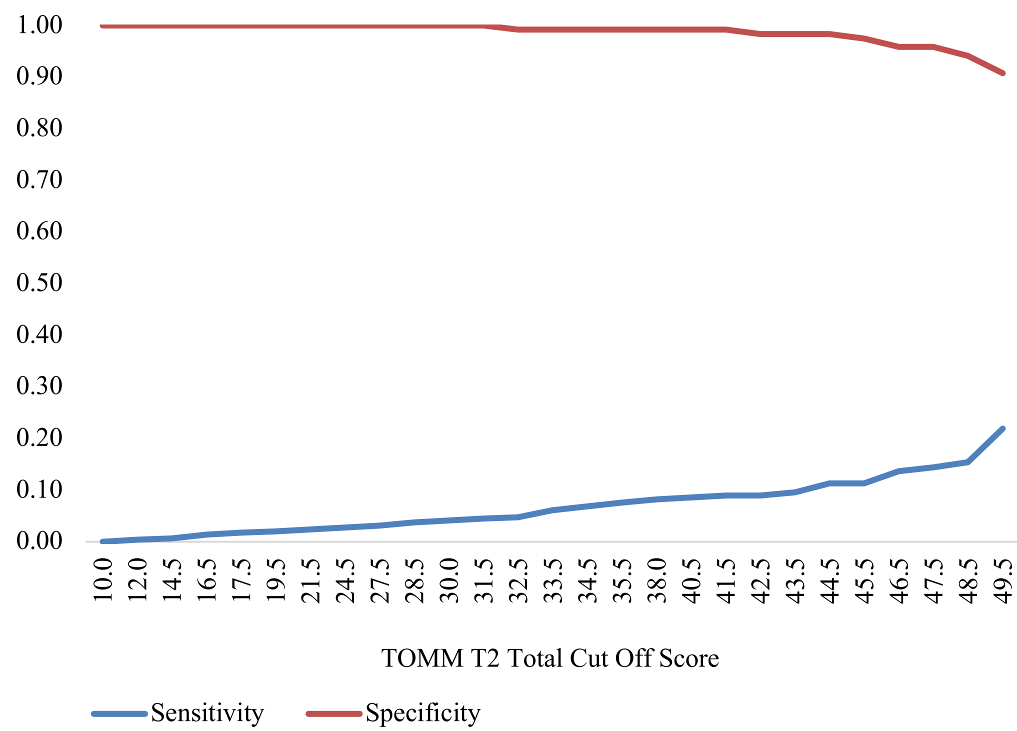

Table 12. *Classification Accuracy Statistics across Observed Scores for the TSI-2-A ATR*

Cut-Score	Sensitivity	Specificity	FPR	PPP	NPP	OCC
3	0.67	0.75	0.25	0.87	0.48	0.69
5	0.54	0.82	0.18	0.88	0.42	0.63
7	0.47	0.93	0.07	0.95	0.42	0.61
9	0.40	0.96	0.04	0.96	0.40	0.56
11	0.32	0.98	0.02	0.98	0.37	0.51
13	0.24	0.98	0.02	0.97	0.34	0.45
15	0.15	0.99	0.01	0.98	0.32	0.39
17	0.10	1.00	0.00	1.00	0.31	0.36
19	0.07	1.00	0.00	1.00	0.30	0.34

Note. Cut-score = observed raw score. Sensitivity = probability of a positive test result given the condition is present (malingering). Specificity = probability of a negative test result, given the condition is absent (not malingering). FPR = False Positive Rate, probability that an individual recieves a positive test result, depite the condition being absent . PPP = Positive Predictive Power, probability the condition is present when the test is positive. NPP = Negative Predictive Power, probability the condition is not present when the test is negative. ACC = Accuracy, proportion of correctly classified individuals to the total sample. All calculations are based on the conditional base rate of 71%.

Table 13. *Classification Accuracy Statistics across Observed Scores for the M-FAST*

Cut-Score	Sensitivity	Specificity	FPR	PPP	NPP	OCC
2	0.86	0.57	0.43	0.83	0.63	0.78
3	0.76	0.72	0.28	0.87	0.55	0.75
4	0.66	0.76	0.24	0.87	0.48	0.69
5	0.58	0.84	0.16	0.90	0.45	0.66
6	0.50	0.85	0.15	0.89	0.41	0.60
7	0.41	0.88	0.12	0.89	0.38	0.55
8	0.35	0.92	0.08	0.91	0.36	0.51
9	0.30	0.92	0.08	0.90	0.35	0.48
10	0.24	0.94	0.06	0.91	0.34	0.44

Note. Cut-score = observed raw score. Sensitivity = probability of a positive test result given the condition is present (malingering). Specificity = probability of a negative test result, given the condition is absent (not malingering). FPR = False Positive Rate, probability that an individual recieves a positive test result, depite the condition being absent . PPP = Positive Predictive Power, probability the condition is present when the test is positive. NPP = Negative Predictive Power, probability the condition is not present when the test is negative. ACC = Accuracy, proportion of correctly classified individuals to the total sample. All calculations are based on the conditional base rate of 71%.

Table 14. *Classification Accuracy Statistics across Observed Scores for the TOMM T2*

Cut-Score	Sensitivity	Specificity	FPR	PPP	NPP	OCC
40	0.08	0.99	0.01	2.18	0.31	0.35
41	0.09	0.99	0.01	0.96	0.31	0.35
42	0.09	0.99	0.01	0.96	0.31	0.35
43	0.09	0.98	0.02	0.93	0.31	0.35
44	0.10	0.98	0.02	0.93	0.31	0.35
45	0.11	0.98	0.02	0.94	0.31	0.36
46	0.11	0.97	0.03	0.92	0.31	0.36
47	0.14	0.96	0.04	0.89	0.31	0.37
48	0.14	0.96	0.04	0.89	0.31	0.38
49	0.15	0.94	0.06	0.87	0.31	0.38
50	0.22	0.91	0.09	0.85	0.32	0.42

Note. Cut-score = observed raw score. Sensitivity = probability of a positive test result given the condition is present (malingering). Specificity = probability of a negative test result, given the condition is absent (not malingering). FPR = False Positive Rate, probability that an individual recieves a positive test result, depite the condition being absent . PPP = Positive Predictive Power, probability the condition is present when the test is positive. NPP = Negative Predictive Power, probability the condition is not present when the test is negative. ACC = Accuracy, proportion of correctly classified individuals to the total sample. All calculations are based on the conditional base rate of 71%.

Figure 15. *Comparison Histograms of Test Total Scores, between Conditions 2 and 4*

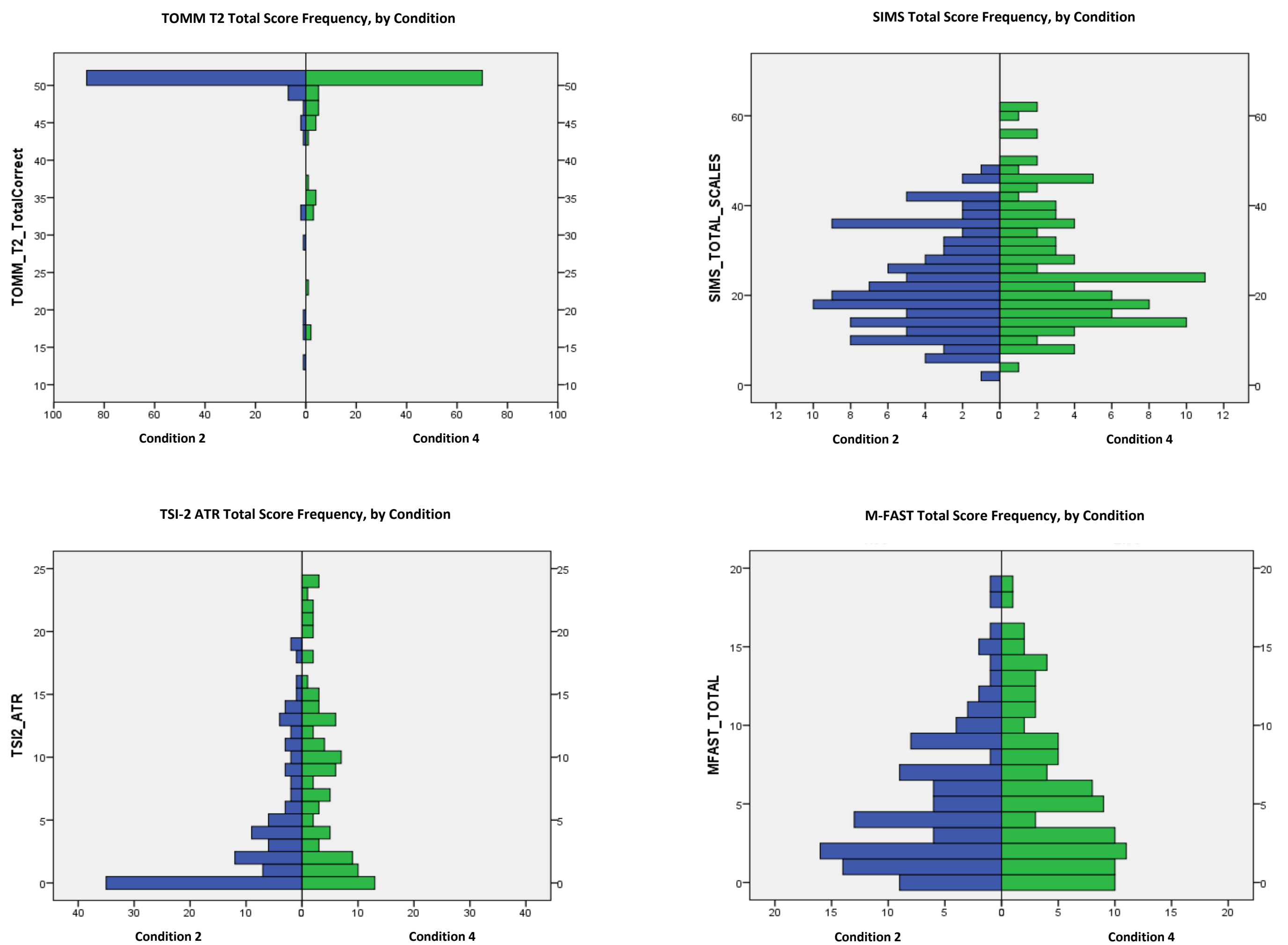

Table 15. *Chi Square of Conditions 2 and 4, Subsample*

	Condition 2		Condition 4			
	n (%)		*n* (%)			
	Pass	Fail	Pass	Fail	X^2	*p*
TSI-2 ATR (Cut-off: ≥15)	37 (100)	0 (0)	49 (81.7)	11 (18.3)	7.65	0.006**
SIMS (Cut-off: >14)	13 (35.1)	24 (64.9)	15 (25)	45 (75)	1.15	0.285
M-FAST (Cut-off: ≥6)	26 (70.3)	11 (29.7)	35 (58.3)	25 (41.7)	1.39	0.237
TOMM T2 (Cut-off: <45)	34 (91.9)	3 (8.1)	51 (85)	9 (15)	1.00	0.317

** indicates p-value significant at the .01 level. * indicates p-value significant at the .05 level.

Condition 2 sample size = 37; Condition 4 = 60.

Table 16. *AUC Summaries by Test based on ROC Curves dichotomized by condition, Subsample*

	AUC	SE	Sig.	95% CI [L, U]
TSI-2 ATR	.689	.056	.002**	.580, .798
SIMS	.597	.059	.109	.482, .712
M-FAST	.597	.059	.110	.481, .712
TOMM T2	.570	.059	.248	.455, .685

** indicates p-value significant at the .01 level. AUC = area under the ROC curve; SE = standard error; Sig. = asymptotic significance, where H0 = 0.5; CI = confidence interval, with Lower and Upper bounds.

Appendices

Appendix A: Pre-Questionnaire

1) Please enter your age (in years).

   a. What is your gender?
   b. Male (0)
   c. Female (1)
   d. Transgender (2)

2) What is your ethnicity/race?
   a. African-American (1)
   b. Hispanic (2)
   c. Caucasian (3)
   d. Asian-American (4)
   e. Other/Multiracial (5) ____________________

3) Do you have a valid driver's license from any state?
   a. Yes (1)
   b. No (2)

4) Are you a registered voter?
   a. Yes (1)
   b. No (2)

5) What is your marital status?
   a. Single but never married (1)
   b. Single but married in the past (2)
   c. Single but living with non-marital partner (3)
   d. Currently married (4)
   e. Widowed (5)

6) Do you have any children?
   a. Yes (1)
   b. No (2)

7) What is your academic year?
   a. Freshman (1)
   b. Sophomore (2)
   c. Junior (3)
   d. Senior (4)
   e. Post-Baccalaureate (Graduated) (5)

8) What is your college major?

9) What is your college grade point average? Please be as specific as possible.

10) In addition to your enrollment, are you currently employed?
   a. No, and I'm not looking (3)
   b. No, but I am looking (4)
   c. Yes, I am employed part-time (2)
   d. Yes, I am employed full-time (1)
   e. Retired (5)

11) If you are employed, what is your current occupation?

12) What are your plans after you receive your degree? (Select which best represents your current plan.)
   a. Pursue additional education (1)
   b. Pursue a new job directly related to my major (2)
   c. Pursue a new job that requires a degree (not related to major area) (3)
   d. Stay with current employer related to my major (6)
   e. Stay with current employer (not related to major area) (4)
   f. Completely undecided (5)

13) What type of career are you planning to pursue once you complete your education?

14) Have you ever served in the military, including the National Guard?
   a. Yes (1)
   b. No (2)

15) Which branch?

16) How long did you serve?

17) Have you ever been involved in active combat?
   a. Yes (1)
   b. No (2)

18) Have you ever been deployed in a war zone?
   a. Yes (1)
   b. No (2)

19) Have you or a family member ever been involved in a lawsuit, either civil or criminal?
   a. Yes (1)
   b. No (2)

20) Who was involved in the lawsuit?

21) How many different lawsuits?

22) What types of lawsuits?

23) What were the outcomes of the case(s)?

24) If you or anyone close to you has ever made any type of claim for monetary damages for emotional or psychological injuries, please provide additional information here:

25) Do you have any beliefs against awarding damages for emotional or psychological injuries in a lawsuit?
   a. Yes (1)
   b. No (2)

26) Please explain your beliefs in regards to awarding damages for emotional or psychological injuries in a lawsuit.

27) Have you ever been involved in a motor vehicle accident that was not your fault in which you sustained physical or psychological injuries?
   a. Yes (1)
   b. No (2)

28) What was the outcome of your experience?
   a. Settled with the other driver (1)
   b. Settled with the other driver's insurance company (2)
   c. Filed a lawsuit and received compensation (3)
   d. Filed a lawsuit and did not receive compensation (6)
   e. Pursued compensation (but no lawsuit) and did not receive it (4)
   f. Did not seek or receive compensation (5)

29) If you filed a lawsuit, what was the outcome of the lawsuit?

30) Have you ever been involved in a motor vehicle accident that was your fault in which you inflicted physical or psychological injuries to the other driver or his/her passenger?
   a. Yes (1)
   b. No (2)

31) What was the outcome of this experience?
   a. Settled with the other driver (1)
   b. The other driver settled with your insurance company (2)
   c. The other driver filed a lawsuit and received compensation (3)
   d. The other driver filed a lawsuit and did not received compensation (6)

e. The other driver pursued compensation (but no lawsuit) and did not receive it (4)
f. Did not seek or receive compensation (5)

32) Please explain the outcome of the lawsuit here:

33) Has anyone close to you ever experienced a motor vehicle accident that was not their fault in which they sustained physical or psychological injuries?
a. Yes (1)
b. No (2)

34) What was the outcome of this experience?
a. Settled with the other driver (1)
b. Settled with the driver's insurance company (2)
c. Filed a lawsuit and received compensation (3)
d. Filed a lawsuit and received no compensation (4)
e. Pursued compensation (but no lawsuit) and did not receive it (6)
f. Did not seek or receive compensation (5)

35) Please explain the outcome of the lawsuit here:

36) Have you ever received mental health treatment services (i.e., psychiatric hospitalization, therapy, counseling, psychiatric medications)?
a. Yes (1)
b. No (2)

37) Please check which mental health treatment services you have received:
a. Counseling (1)
b. Therapy (2)
c. Psychiatric Hospitalization (3)
d. Psychiatric Medications (4)
e. Other (5) ____________________

38) Please describe the reasons you received these services.

39) Have you ever participated in a mental health evaluation where you were administered psychological tests?
a. Yes, for my evaluation (1)
b. Yes, as a research volunteer (3)
c. No (2)

40) Please explain why psychological tests were administered. (If you don't know the names, please describe them briefly.)

41) Please list the tests you were administered. (If you don't know the names, please describe them briefly.)

42) Have you ever received a psychiatric/psychological diagnosis?
  a. Yes (1)
  b. No (2)

43) Please list all psychiatric/psychological diagnoses.

44) Do you experience or have you been diagnosed with any medical/health problems?
  a. Yes (1)
  b. No (2)

45) Please describe your medical/health problems here:

46) Have you ever worked in the mental health services area?
  a. Yes (1)
  b. No (2)

47) Please list your past or current mental health services employment:

48) Has anyone close to you ever worked in the mental health services area?
  a. Yes (1)
  b. No (2)

49) Who worked in the mental health services area?

50) Please list the type of mental health services employment someone close to you was engaged in.

51) Have you or anyone close to you ever worked in the legal field?
  a. Yes (1)
  b. No (2)

52) Who worked in the legal field and what type of employment did you/they have?

Appendix B: Instructional Conditions/ Case Scenarios

Condition 1 (Post-litigation, no suggestion)

Imagine you were the plaintiff in a civil lawsuit against a major wholesaler. You sued for damages related to a car accident you were in. On the day of the accident you were driving home from school, when you were involved in a collision with a major wholesaler's delivery truck. The driver of the truck had run a red light, and was clearly liable. The truck hit the rear side of your car causing you to be thrown against your seatbelt with great force. Your car was totaled. You sustained scrapes, bruises, and a bruised rib. You were taken by ambulance to the hospital, observed, and released later the same day. The company acknowledged liability, replaced your vehicle and paid for any miscellaneous costs you incurred. In addition, as a result of your civil suit you were awarded monetary damages that you felt were satisfactory.

After the accident you initially experienced a great deal of discomfort and pain due to the bruised rib, but the pain went away within a couple of weeks. You did miss a couple days of work, but have been compensated for those. However, you are still experiencing emotional difficulties including jumpiness/nervousness while driving, avoidance of the location of the accident (even though it is the quickest way for you to get home), avoiding conversations about the accident, having bad dreams about the accident, and you have an exaggerated startle response. You are also experiencing difficulty with concentration, trouble remembering things, and you feel "foggy." You spoke about these symptoms to your physician, who requested that you complete the following assessments in order to determine your level of impairment.

Now you will complete an evaluation with a member of the research staff. Please respond to the following questions as if you are the person in this situation.

Condition 2 (No litigation, no suggestion)

Imagine you were in a motor vehicle accident. On the day of the accident you were driving home from school, when you were involved in a collision with a major wholesaler's delivery truck. The driver of the truck had run a red light, and was clearly liable. The truck hit the rear side of your car causing you to be thrown against your seatbelt with great force. Your car was totaled. You sustained scrapes, bruises, and a bruised rib. You were taken by ambulance to the hospital, observed, and released later the same day. The company has acknowledged liability, has replaced your vehicle and paid for any miscellaneous costs you incurred.

After the accident you initially experienced a great deal of discomfort and pain due to the bruised rib, but the pain went away within a couple of weeks. You did miss a couple days of work, but have been compensated for those. However, you are still experiencing emotional difficulties including jumpiness/nervousness while driving, avoidance of the location of the accident (even though it is the quickest way for you to get home), avoiding conversations about the accident, having bad dreams about the accident, and you have an exaggerated startle response. You are also experiencing difficulty with concentration, trouble remembering things, and you feel "foggy." You spoke about these symptoms to your physician, who requested that you complete the following assessments in order to determine your level of impairment.

Now you will complete an evaluation with a member of the research staff. Please respond to the following questions as if you are the person in this situation.

Condition 3 (Active litigation, no suggestion)

Imagine you are the plaintiff in a civil lawsuit against a major wholesaler. You are suing for damages related to a car accident you were in recently. On the day of the accident you were driving home from school, when you were involved in a collision with a major wholesaler's delivery truck. The driver of the truck had run a red light, and was clearly liable. The truck hit the rear side of your car causing you to be thrown against your seatbelt with great force. Your car was totaled. You sustained scrapes, bruises, and a bruised rib. You were taken by ambulance to the hospital, observed, and released later the same day. The company has acknowledged liability, has replaced your vehicle and paid for any miscellaneous costs you incurred.

After the accident you initially experienced a great deal of discomfort and pain due to the bruised rib, but the pain went away within a couple of weeks. You did miss a couple days of work, but have been compensated for those. However, you are still experiencing emotional difficulties including jumpiness/nervousness while driving, avoidance of the location of the accident (even though it is the quickest way for you to get home), avoiding conversations about the accident, having bad dreams about the accident, and you have an exaggerated startle response. You are also experiencing difficulty with concentration, trouble remembering things, and you feel "foggy." You spoke about these symptoms to your physician, who suggested you contact a lawyer. You decided to contact a well-known injury lawyer, and together, decided to proceed with litigation alleging sustained emotional difficulties and post-traumatic stress disorder from the motor vehicle accident. Your lawyer explains to you that the wholesaler's legal team is requesting that you complete a number of assessments to determine your level of impairment.

Now you will complete an evaluation with a member of the research staff. Please respond to the following questions as if you are the person in this situation.

Condition 4 (Active litigation, suggestion)

Imagine you are the plaintiff in a civil lawsuit against a major wholesaler. You are suing for damages related to a car accident you were in recently. On the day of the accident you were driving home from school, when you were involved in a collision with a major wholesaler's delivery truck. The driver of the truck had run a red light, and was clearly liable. The truck hit the rear side of your car causing you to be thrown against your seatbelt with great force. Your car was totaled. You sustained scrapes, bruises, and a bruised rib. You were taken by ambulance to the hospital, observed, and released later the same day. The company has acknowledged liability, replaced your vehicle and paid for any miscellaneous costs you incurred.

After the accident you initially experienced a great deal of discomfort and pain due to the bruised rib, but the pain went away within a couple of weeks. You did miss a couple days of work, but have been compensated for those. However, you are still experiencing emotional difficulties including jumpiness/nervousness while driving, avoidance of the location of the accident (even though it is the quickest way for you to get home), avoiding conversations about the accident, having bad dreams about the accident, and you have an exaggerated startle response. You are also experiencing difficulty with concentration, trouble remembering things, and you feel "foggy." You spoke about these symptoms to your physician, who suggested you contact a lawyer. You decided to contact a well-known injury lawyer, and together, decided to proceed with litigation alleging sustained emotional difficulties and post-traumatic stress disorder from the motor vehicle accident. Your lawyer explains to you that the wholesaler's legal team is requesting that you complete a number of assessments to determine your level of impairment. Your lawyer also tells you that the more impaired you appear on the following assessments, the higher amount of monetary damages you will be awarded.

Now you will complete an evaluation with a member of the research staff. Please respond to the following questions as if you are the person in this situation.

Appendix C: Post-questionnaire

In civil litigation, damages are often awarded to return the injured person to his or her level of functioning prior to the incident (e.g., car accident). The law allows an injured party to seek both pecuniary and non-pecuniary damages to redress for injuries sustained. Pecuniary damages provide compensation for direct financial loss sustained from the injury, such as lost days at work, damage to a vehicle, or costs incurred. Non-pecuniary damages compensate the injured party for non-tangible losses such as pain and suffering or emotional damages. In the present case, the major wholesaler has already paid for your pecuniary damages (i.e., lost days at work, damage to your vehicle, costs incurred). The company has not paid for your pecuniary damages (i.e., pain and suffering or emotional damages).

1) Which of the following will pecuniary damages compensate a litigant for?
    a. Lost days at work (1)
    b. Pain and suffering (2)
    c. Emotional Damages (3)

2) Which of the following will non-pecuniary NOT damages compensate a litigant for?
    a. Pain and suffering (1)
    b. Lost days at work (2)
    c. Emotional Damages (3)

3) Which type of damages has the company already paid you for in this scenario?
    a. Pecuniary (e.g., lost days at work, damage to your vehicle, costs you incurred) (1)
    b. Non-pecuniary (e.g., pain and suffering and emotional damages) (2)

4) What sum of money, if any, would fairly and reasonably compensate you for your emotional damages (i.e., non-pecuniary damages) that resulted from the accident portrayed in the scenario? Do *not* include compensation for pecuniary damages (e.g., lost days at work, damage to your vehicle, costs you incurred).

5) Please detail what percentage you considered each item, if at all, in response to questions you were asked about the scenario. (The total of all items cannot be more than 100%.)
    a. _______ Emotional Symptoms (1)
    b. ______ Medical Costs (includes costs for outpatient therapy and medication) (2)
    c. ______ Cognitive Symptoms (3)
    d. ______ Loss of Future Earnings (4)
    e. ______ Defendant's Behavior (Major Wholesaler) (5)
    f. ______ Loss of Pleasure from Life (6)
    g. ______ Other (7)
    h. ______ Other (8)

6) To what extent do you believe that the accident was a relevant factor in creating the emotional difficulties you are hypothetically experiencing?
   a. Most relevant factor (1)
   b. Very relevant, but not the most relevant factor (2)
   c. Somewhat relevant (3)
   d. Not at all relevant (4)

7) Please rate the severity of your hypothetical injuries on a scale of 1 (no distress/disability) to 10 (severe distress/disability):

	No distress/ disability (1)	(2)	(3)	(4)	Moderate Distress/ Disability (5)	(6)	(7)	(8)	(9)	Severe distress/ disability (10)
Cognitive Functioning (1)										
Social Functioning (2)										
Psychological Functioning (3)										
Occupational/ Academic Functioning (4)										
Emotional Functioning (5)										

8) How much did you exaggerate the symptoms you were experiencing in the forensic portion of the assessment, if at all?
   a. Not at all (1)
   b. A Little Bit (2)
   c. Moderately (3)
   d. Quite a Bit (4)
   e. Extremely (5)

9) How well do you think you exaggerated the following types of symptoms in the forensic portion of the assessment?

	Not at all	A Little Bit	Moderately	Quite a bit	Extremely
Cognitive Functioning (1)					
Social Functioning (2)					
Psychological Functioning (3)					
Occupational/Academic Functioning (4)					
Emotional Functioning (5)					

10) In the scenario, did anyone suggest that you would receive more compensation if your injuries were greater?
   a. Yes (1)
   b. No (2)

11) In the scenario, what was the purpose for which you completed the evaluation?

   a. Physician's request (1)
   b. Litigation purposes (2)

12) Did you believe you would receive more compensation if your injuries were greater?
   a. Yes (1)
   b. No (2)

13) Throughout the evaluation, how consistently do you believe you followed the instructions given in your scenario (0 equals 0 percent of the time and 100 equals 100 percent of the time)? [Slide bar to indicate how consistently you followed instructions during the evaluation.]

14) Are there ways for assessment measures to detect if people are exaggerating their symptoms?
   a. Yes (1)
   b. No (2)

15) Briefly describe your impressions of the purpose or goals of this study: