The Assessment and Detection Feigned Symptoms that may persist after a Mild Traumatic Brain Injury: An Analogue Investigation

By

Ms. Sonya Dhillon, H.BSc.

A thesis submitted in conformity with the requirements for the degree of Master or Arts

Graduate Department of Psychological Clinical Science

University of Toronto

© Copyright by Sonya Dhillon 2017
The Assessment and Detection Feigned Symptoms that can Persist after a Mild Traumatic Brain Injury: An Analogue Investigation

Sonya Dhillon

Master of Arts

Graduate Department of Psychological Clinical Science

University of Toronto

2017

Abstract

Psychologists have many instruments to assess the over-reporting of symptoms that may persist after a mild traumatic brain injury (mTBI), including symptom validity tests (SVTs) and performance validity tests (PVTs). The comparative predictive capacity of SVTs “embedded” in the Minnesota Multiphasic Personality Inventory 2 Restructured Form (MMPI-2-RF) and the Personality Assessment Inventory (PAI) versus three PVTs were examined in a simulation study. Participants were administered the MMPI-2-RF, PAI and PVTs and instructed to either feign symptoms or respond honestly. Using a series of hierarchical logistic regression analyses, the performance of SVTs and PVTs to differentiate feigners from honest responders was examined. The Response Bias Scale (RBS) of the MMPI-2-RF was the best predictor relative to all other SVTs and PVTs. Those assessing the possible feigning of mTBI would be well served by including the MMPI-2-RF in their psychological test battery.
Acknowledgements

There are many individuals that I owe my gratitude to.

To my graduate supervisors, Dr. Bagby and Dr. Zakzanis. I have been so fortunate to learn from not one, but two “rockstars” in the clinical sciences. I am grateful to you both, for providing me with the necessary guidance to complete this ambitious project. Thank you for bringing meaning to the research, and helping me get out of my own head when I needed to. Thank you for believing in me before I did. I am so lucky to have your mentorship. To my committee member, the clinical faculty, and the administrative staff in my department; you all have played an important role in my professional life, and have helped me attain this goal.

To my cohort and fellow students: thank you for the support over the past two years. Simply put, there is no greater feeling than knowing that you were all there for me when I needed you. To my lab family, the research assistants, post doc extraordinaire, Dr. Kushner, for their tireless efforts in helping me develop and execute this project. Shauna, Tara, Allison, Angie and Nicole; thank you for being the sisters that I never had.

To the love of my life, Phil. Who worked harder to achieve this monumental milestone; you or I? You. You have worked so hard to ensure that life outside of grad school was easy and full of happiness. Thank for all that you do for our family. Thank you for loving me the way that you do. Thank you for making me a better version of myself every day and believing that I can “save the world.”

To my dear parents, mama and daddy: thank you for sacrificing your dreams to allow me to follow mine. I am eternally grateful for guidance that you have both provided me throughout my life. Thank you for your unconditional love, and support for your “head in the clouds” daughter. You taught me that the sky is not the limit, and that I am in charge of my success.
Thank you for giving me the tools to thrive in life.

To my late brother, Ryan. This moment is bittersweet. I am so happy that I have accomplished this goal that I know would make you so proud, yet it pains me that I cannot share it with you. Thank you for always believing in me. I am so fortunate to have had a brother like you who worked tirelessly to ensure his baby sister grew up to be a strong, confident woman.

Thank you for being the motivation behind my decision to pursue this career.

Thank you for teaching me that life is not measured by the number of breaths we take, but by the moments that take our breath away.
# Table of Contents

Acknowledgements iii

Table of Contents v

List of Tables, Figures, and Appendices. vi

Introduction 1

Over-reporting 2

Over-reporting of Symptoms that May Persist after a Mild Traumatic Brain Injury Symptoms. 2

Performance Validity Tests. 5

Symptom Validity Tests 8

Utility of PVTs and SVTs 13

The Current Study 13

Hypotheses. 14

Method. 14

Participants. 15

Measures. 15

Procedure. 18

Statistical Analyses. 21

Results. 22

Discussion. 28

References. 35
List of Tables

Table 1. Participant self-reported ethnicity. 47
Table 2. Means, standard deviations, and within group comparisons of standard instruction condition group at time 1 and time 2. 48
Table 3. Means, standard deviations, F tests, and effect size estimates between groups at time 1. 49
Table 4. Means, standard deviations, and between group comparisons of standard instruction (SI) and over-reporting instruction (OR) groups. 50
Table 5. Results of hierarchical logistic regression analyses, model characteristics and diagnostic efficiency statistics. 51
Table 6. Hierarchical Logistic Regression Analysis comparing the MMPI-2-RF and PAI scales and indices differentiating SI from OR groups. 52
Table 7. Hierarchical Logistic Regression Analysis comparing the PVT and MMPI-2-RF scales in differentiating SI from OR groups. 53
Table 8. Hierarchical Logistic Regression Analysis comparing the PVTs and PAI scales and indices in differentiating SI from OR groups. 54

List of Figures

Figure 1. Protocol Elimination. 19
Figure 2. Design and Procedure. 20

List of Appendices

Appendix A. Post Experimental Questionnaire. 55
Appendix B. mTBI and related symptoms description provided to over-reporting instruction participants. 56
Appendix C. Case vignette and simulation instructions provided to over-reporting participants. 57
The Assessment and Detection Feigned Symptoms that can Persist after a Mild Traumatic Brain Injury: An Analogue Investigation

According to the Canadian Institute for Health Information (CIHI; 2007) the direct cost of traumatic brain injury (TBI) in Canada in 2000–2001 was $151.7 million. Mild TBI (mTBI) constitutes 70–90% of all TBIs, with annual rates of hospital treatment ranging from 100 to 300 per 100,000 (Cassidy et al., 2004). The diagnosis of mTBI is based on the acute injury characteristics coupled with the absence of intracranial injury on neuroimaging examination. Acute injury characteristics include whether consciousness was lost, and for how long; duration of posttraumatic amnesia and on the results on the Glasgow Coma Scale (American Congress of Rehabilitation Medicine [ACRM]; 1993). The symptoms that can persist after a mTBI are varied and include cognitive (e.g., memory impairments, reduced concentration and attention), psychological (e.g., low mood, irritability) and somatic (e.g., headaches, body aches) concerns. The breadth, severity and veracity of such symptoms are usually examined in the context of a neuropsychological assessment, which typically includes clinical interviewing and psychometric testing after the injury was sustained. Neuropsychologists rely on a patient’s self-report and file information to assess whether or not a mTBI may have occurred.

The expected recovery period following an mTBI is one to three months (Frencham, Fox, & Mayberry, 2005; Mogge, Lepage, Bell & Ragatz, 2010; Pertab, James & Bigler, 2009; Rohling et al., 2011); however, 5–15% of individuals with mTBI report symptoms, and in some cases, functional impairment, for up to a year after this expected recovery period (Carroll et al., 2004; Mogge, Lepage & Ragatz, 2010; Ponsford, 2005). The persistence of symptoms may be attributed to different factors, such as external incentives for an examinee post injury (Bianchini, Curtis & Greve, 2006) or the perception of injury following the label of mTBI (Rettmann,
Over-reporting

In neuropsychological assessment contexts, in which the symptoms that can persist after an mTBI are being evaluated, individuals may over-report their symptoms for some external incentive (e.g., monetary compensation, time off work; Bianchini, Curtis & Greve, 2006). If this over-reporting response style is excessive, the results obtained from a neuropsychological or psychological test may lead to an inaccurate clinical appraisal (Sellbom & Bagby, 2008; 2010). Test outcomes that are due to over-reporting can also compromise the validity of the test (Dhillon et al., 2017). Based on the clinician’s interpretation of these altered test results, one may infer that an examinee is impaired when they are not (An, Zakzanis & Joordens, 2012). In recognition of this issue, assessing for over-reporting is an important step in a neuropsychological assessment (see, for e.g., Berry, Baer, & Harris, 1991; Franzen, Iverson, McCracken, 1990; Rogers, Harrell, & Liff, 1993; Schretlen, 1988).

Over-reporting of Symptoms that May Persist after a Mild Traumatic Brain Injury

Symptoms

The over-reporting of the symptoms that may persist after an mTBI (e.g., cognitive, psychological, somatic) is present in a significant number of examinees referred for neuropsychological assessments in scenarios where injuries are potentially compensable (Iverson, 2005; Pertab, James, & Bigler, 2009; Zielinski, 1994). Including tests that assess over-reporting in neuropsychological assessments that assess for over-reporting of these symptoms can maximize the level of confidence in test results, and hence, improve clinical decision-making (Faust, Guilmette & Arkes, 1988).

In neuropsychological assessments, there are different methods to assess over-reporting.
These methods include clinical judgment and a variety of psychometric tools. The validity of these two methods has been examined extensively in the behavioral sciences literature and conclude rather definitively that clinical judgment alone is inadequate to detect accurately over-reporting (Faust, Guilmette & Arkes, 1988; Heaton, Smith, Lehman & Vogt, 1978; Slick, Strauss & Hultsch, 2004; van Gorp et al., 1999). In their seminal paper, Dawes, Faust and Meehl (1989) compared clinical judgment and empirically based methods (e.g., psychological tests) in differential diagnoses. Empirically based methods (e.g., objective tests), in general, were superior across settings compared to clinical judgment (see for e.g., Grove et al., 2000; Leli & Filskov, 1984).

The value of including both neuropsychological and psychological tests with well-established and validated “validity” indices has been endorsed by the Institute of Medicine of the National Academics (IOM, 2015). The IOM suggests that validity indices improve accuracy with respect to false positives and false negatives. Performance validity tests (PVTs) and symptom validity tests (SVTs) are two categories of tests designed to measure over-reporting.

PVTs are performance-based cognitive tests that measure whether an examinee performs in a manner that is consistent with their “actual” level of ability; that is did the examinee put forth maximal effort on a given test(s) (Bush et al., 2005). For most PVTs, a “floor effect” detection strategy is used; if performance falls below chance to a statistically significant degree this indicates invalid performance (i.e., evidence of over-reporting). The cognitive capacities that are necessary to successfully pass PVTs usually remain intact and are resistant to injury (Slick et al., 2003; Teichner & Wagner, 2004) and as such, scores at or below chance are rare. A clinician makes conclusions about the validity of an examinee’s performance based on established cut-off scores.
SVTs are symptom-based tests that assess whether an examinee is over-reporting in a manner that is inconsistent relative to a patient or normative groups. SVTs are used to establish the probability that a given test result is affected by a tendency to over-report symptoms (Bush et al., 2005; Rohling et al., 2011; Victor, Boone, Serpa, Buehler, & Ziegler, 2009).

Self-report instruments can include embedded SVTs. Embedded SVT scales and indices assess the validity of test performance and are “built in” to an existing test. SVT scales and indices (i.e., over-reporting scales) employ strategies to identify those who are reporting their symptoms non-credibly. Elevations on these indicators are typically rare in community samples, but are not uncommon in civil-legal samples (Griffin, Normington, May, & Glassmire, 1996; Lees-Haley, 1991; 1997). Like other psychometric tools (see for e.g., Clark & Watson, 1995), these SVT scales and indices are often cross-validated across multiple samples, demonstrating that high scores on these scales are indeed uncommon in multiple populations, ranging from those with severe psychopathology to community samples.

Previous studies have shown that a distinction can be made between performance validity and symptom validity and over-reporting of cognitive impairment and psychological symptoms, respectively (Grossi et al., 2017; Nelson et al., 2007; Ruocco et al., 2008; Van Dyke, Millis, Axelrod, & Hanks, 2013). In most cases, tests that are sensitive to measuring over-reporting of cognitive complaints are not sensitive to over-reporting emotional/psychological complaints and vice versa (Grossi et al., 2017; Nelson, Sweet, Berry, Bryant & Granacher, 2007; Ruocco et al., 2008; Sharf et al. 2017). For example, the scores obtained of the Test of Memory Malingering (TOMM; Tombaugh, 1996) are not impacted by depression (Rees, Tombaugh & Boulay, 2001), but are impacted by exaggerated or deliberately faked memory impairment in clinical situations.

Given that self-reported cognitive, emotional and psychological complaints are common
symptoms that can persist after an mTBI, the assessment for the presence of these symptoms may rely on both PVTs and SVTs to improve accuracy in their when assessing for over-reporting of these symptoms. With respect to practical application, PVTs are often used in neuropsychological settings whereas SVTs have been used across settings (see, for e.g., Ruocco et al., 2008).

**Performance Validity Tests**

Although many different PVTs have been developed, only the PVTs that are used as independent variables (IV) in the current study, and those that formed the Response Bias Scale (RBS; Gervais et al, 2007), which is also an IV (see symptom validity test section below) are reviewed. According to a review by Larrabee (2012), the failure on three out of three PVTs represents specificity and sensitivity rates above 99% (Larrabee, 2012). As such, it is recommended to administer three PVTs to maximize confidence in clinical decision making with respect to the credibility of tests results obtained.

The Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997), Test of Memory Malingering (TOMM; Tombaugh, 1996) and the Word Memory Test (WMT; Green, Allen, & Astner, 1996) are the three PVTs that were used in the initial development and validation of the RBS.

The Rey Fifteen Item Test (Rey-FIT; Rey, 1964), the Reliable Digit Span (RDS; Greiffenstein, Baker, & Gola, 1994) and the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997) are PVTs that are commonly used in neuropsychological assessments (Larrabee, 2012; Lezak, 2004). Studies have shown that these PVTs are valid and reliable in measuring performance validity in multiple neuropsychological assessment contexts (e.g., disability evaluations, neuropsychological assessment) (Greher & Wodushek, 2017;

**Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997).** The CARB is a forced choice digit recognition test that assesses possible over-reporting of memory impairment. The literature has shown that the CARB has acceptable diagnostic efficiency values in severe and moderate mTBI populations (Allen et al., 1997; Green, Gervais, Astner, Kiss, & Allen, 1993) and these results have been replicated (Green, Rohling, Lees-Haley, & Allen, 2001). In a simulation study, those that were instructed to simulate cognitive impairment had significantly different scores than those that performed to the best of their abilities and a patient sample (Conder et al., 1992).

**Test of Memory Malingering (TOMM; Tombaugh, 1996).** The TOMM is a forced choice object recognition test that assesses possible over-reporting of memory impairment. Multiple studies conducted by Rees, Tombaugh, Gansler and Moczynski (1998) demonstrated that the TOMM distinguished suspected malingers and simulators from those who were putting forth maximum effort. The TOMM has demonstrated strong sensitivity and specificity levels in multiple populations (e.g., university students, patients with TBI, and hospital outpatients). In addition, these indicators of diagnostic efficiency have been obtained in simulation and known groups deigns.

**Word Memory Test (WMT; Green, Allen, & Astner, 1996).** The WMT is a computer-based test that is designed to measure both verbal memory and over-reporting of memory symptoms. The WMT has been shown to identify possible over-reporting on memory tasks (Green, 2005; Green, Montijo, & Brockhaus, 2011) and is performs similarly with its psychometric properties (e.g., sensitivity and specificity) to other PVTs, such as the TOMM (Bauer, O'Bryant, Lynch, McCaffrey, & Fisher, 2007; Heyanka et al., 2015; Sollman & Berry,
Rey Fifteen Item Test (Rey-FIT; Rey, 1964). The Rey-FIT is a performance validity test that measures possible over-reporting of memory complaints. Administration details are provided in the method section. The Rey-FIT has acceptable sensitivity and specificity in a variety of settings (e.g., neuropsychological, forensic, disability, psychiatric and research; Reznek, 2005; Vickery et al., 2001). The sensitivity of the test increases when a recognition trial is added to the original form. In the recognition trial, an examinee is instructed to select stimuli that they had seen before in the first trial from a group of stimuli that include foils. The specificity of the original form is maintained when the recognition trial is included (Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002; Reznek, 2005).

Reliable Digit Span (RDS; Wechsler, 1987). The RDS is a measure derived from the Digit Span subtest of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) and Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987). Administration details are provided in the method section. This embedded PVT has acceptable sensitivity and specificity in different settings, including brain injured populations (Greiffenstein, Gola, & Baker, 1995; Meyers & Volbrecht, 1998), forensic (Duncan & Ausborn, 2002), and simulation studies (Iverson & Franzen, 1996).

Victoria Symptom Validity Test (VSVT; Slick et al., 1997). The VSVT assesses possible over-reporting of cognitive impairments. Administration details are provided in the method section. The VSVT has strong sensitivity and specificity rates in a variety of settings similar to the Rey-FIT (Frazier, Youngstrom, Naugle, Haggerty, & Busch, 2007; Grote et al., 2000; Slick et al., 2003; Ruocco, 2016; Strauss et al., 2002).
Symptom Validity Tests

Two omnibus measures of personality and psychopathology that include embedded SVT scales and indices are the Minnesota Multiphasic Personality Inventory-2-RF (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011) and the Personality Assessment Inventory (PAI; Morey, 1991). Many of these SVT scales and indices measure over-reporting of symptoms that can persist after an mTBI – cognitive impairment (Rohling et al, 2011), emotional problems (Mathias & Coats, 1999) and somatic complaints (Nampiaparampil, 2008).

*Minnesota Multiphasic Personality Inventory 2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008/2011).* The MMPI-2-RF is comprised of 338 of the original 567 items of the MMPI-2 (Butcher et al., 2001). The MMPI-2-RF includes six sets of scales: Validity, Higher-Order, Restructured Clinical, Specific Problem, Interest, and Personality Psychopathology scales. The validity scales are designed to measure various aspects of protocol validity: the number of items left unanswered, non-content based invalid responding (e.g., nay saying, yay-saying, random responding), under-reporting, and over-reporting. The current study focused on the validity of the MMPI–2-RF scales designed to measure over-reporting of symptoms that can persist after an mTBI. Some of the over-reporting validity scales were designed to measure over-reporting of general psychopathology such as psychotic or depressive symptoms (Ben-Porath & Tellegen, 2008/2011); other scales are designed to measure over-reporting of signs and symptoms that are specific to disability evaluations, such as depression, anxiety, pain, multiple health systems, and cognitive impairment (Ben-Porath & Tellegen, 2008/2011; Gervais, Ben-Porath, Wygant, & Green, 2007; Lees-Haley, English, & Glenn, 1991). The MMPI-2-RF SVT scales can therefore be useful when assessing over-reporting of symptoms that can persist after an mTBI.
The Infrequent Responses Scale (F-r). The F-r is comprised of 32 items that were rarely endorsed in the MMPI-2-RF normative sample (i.e., were answered in the keyed direction by 10% or less). F-r has been identified as a general indicator that is sensitive to over-reporting of psychological, cognitive and somatic symptoms (Ben-Porath, 2012; Ben-Porath & Tellegen, 2008/2011; Sellbom, Toomey, Wygant, Kucharski, & Duncan, 2010).

The Infrequency Psychopathology Responses scale (Fp-r). The Fp-r is comprised of 21 items, 17 of which were included on the MMPI-2 version of the scale (Fp). Four additional items were added to Fp-r based on multiple regression analyses indicated incremental validity of the scale (Tellegen & Ben-Porath, 2008). Fp-r has been identified as an indicator of over-reported psychopathology, and is effective in differentiating bonafide psychiatric patients from over-reporters (Ben-Porath, 2013). In a “known-groups” design that used the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) to classify groups as malingering or not, F-r and Fp-r were the two over-reporting scales of the MMPI-2-RF that best differentiated the two groups (Sellbom, Toomey, Wygant, and Kucharsli, 2010).

The Infrequent Somatic Response Scale (Fs). The Fs scale is comprised of 16 items and was developed by Wygant, Ben-Porath and Arbisi (2004). Items with somatic content that were endorsed by 25% of the patients in two medical samples and one chronic pain sample (N = 55000) were included in the scale. Fs was designed to detect the over-reporting of somatic symptoms, including somatic complaints rarely endorsed by medical and chronic pain patients (Ben-Porath, 2013). Wygant (2007) examined the scale in a simulation, known-groups, and mental health samples, and found that Fs was elevated among patients who also failed PVTs. Fs scores were found to be significantly higher in participants instructed to simulate head injury symptoms.
The Symptom Validity Scale-revised (FBS-r; Ben-Porath & Tellegen, 2008). The FBS-r is comprised of 30 of the 43 items that comprised the original FBS – from the MMPI-2. The three “infrequency” scales, F-r, Fp-r, and Fs, do not overlap in content. FBS-r, however, shares three items with Fs and one with Fp-r. The FBS-r scale was developed to address the lack of scales on the MMPI-2 to measure the over-reporting style in personal injury litigants. To this end, the FBS was constructed, utilizing a rationale approach, by identifying item content and comparing responses of personal injury litigants that were judged to be malingers and those that were not. Item selection was informed with the assumptions that that personal injury litigants would put forth effort to appear honest and psychologically healthy, aside from the impact of the injury, but also avoid admission of preexisting psychological symptoms (Lees-Haley et al., 1991). The FBS-r was developed to measure over-reported somatic and cognitive complaints (Ben-Porath & Tellegen, 2008). The literature has shown that the FBS-r (Lees-Haley, English, & Glenn, 1991) and Fs (Ben-Porath & Tellegen, 2008/2011) are valid indicators of over-reporting in mTBI-related assessments (Ben-Porath, 2013).

The Response Bias Scale (RBS; Gervais, Ben-Porath, Wygant, and Green, 2007). The RBS is comprised of 28 items, which were identified through a series of multiple regression analyses. Items that were identified as significant predictors of failure on the three PVTs (discussed previously) were retained. The RBS shares four items with F-r, two with Fp-r and Fs, and four on FBS-r. The RBS was designed to assess feigned cognitive impairment and been found to be a significant predictor of exaggerated memory impairment in a simulation study (Sullivan & Elliot, 2012). In addition, the RBS provides incremental validity over three PVTs in the assessment of exaggerated memory complaints (Gervais, Wygant, Ben-Porath & Green, 2007; Gervais et al., 2008). The RBS was constructed by identifying sets of items of the MMPI-
2-RF that were associated with the failure on two out of three PVTs – the Test of Memory Malingering (Tombaugh, 1996), the Word Memory Test (Green, Astner & Allen, 1997), and the Computerized Assessment of Response Bias (Allen, Conder, Green, & Cox, 1997).

In a study that employed a simulation and known groups design, results show that the MMPI-2-RF over-reporting validity scales were effective at detecting symptom over-reporting in civil forensic settings (Wygant et al., 2008) and in a subsequent known groups study examining feigned mental disorders and cognitive impairments (Rogers, et al., 2011).

**Personality Assessment Inventory (PAI; Morey, 1991).** The PAI includes 22 non-overlapping sets of scales and an additional 10 conceptually derived subscales or indices: clinical, treatment consideration, interpersonal and validity. The validity scales and indices measure similar constructs to the MMPI-2-RF validity scales such as inconsistent responding, under-reporting, and over-reporting. The current investigation focused on the validity of the PAI scales and indices designed to measure symptom over-reporting (Morey, 1991).

**Negative Impression Management (NIM; Morey, 1991).** The NIM scale is comprised of eight items that do not overlap with any other scale or index on the PAI. NIM uses assessment of rarely endorsed symptoms. The NIM scale was designed to detect exaggerated presentation of bizarre and/or unlikely symptoms.

**Rogers Discriminant Function (RDF; Rogers, Sewell, Morey & Ustad, 1996).** The RDF is derived from a combination of discriminant function weighted scores from various PAI scales and was designed to detect responses of patterns inconsistent with clinical populations based on a discriminate function analysis using 246 feigners and 221 patients. Morey and Lanier (1998), and Bagby et al. (2002), both found support for the RDF using simulation studies with undergraduate students.
**Malingering Index (MAL).** The MAL index is derived from eight configural features of various PAI scales and was designed to detect over- and under-endorsed items inconsistent with clinical populations (Morey, 1996). MAL has been found to be effective in identifying response styles in which respondents over-endorse items rarely endorsed in clinical samples (Morey, 1996), but has also been unable to discriminate psychiatric patients from research participants who were asked to malingering (Bagby et al., 2002).

In a recent simulation study investigating the performance of the PAI SVTs in a mTBI context, Keiski, Shore, Hamilton, and Malec (2015) found that the SVT indices of the PAI (i.e., NIM, RDF and MAL) were sensitive to simulated symptoms (e.g., memory impairment, headaches, somatic), in the context of TBI.

**Negative Distortion Scale (NDS; Mogge, Lepage, Bell & Regatz, 2010).** The NDS is comprised of 15 items. Items selected were 1.28 standard deviations (the lowest 10%) below the mean average item scale of the relevant scale in the development sample. The NDS was designed to measure over-reporting in a population where the level of psychopathology is low, and has been shown to be sensitive in identifying over-reporting in these contexts (Mogge et al., 2010). In addition, NDS scores were significantly higher in a group of over-reporters as identified by the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992).

**Malingered Pain-Related Disability-Discriminant Function (MPRD-DF; Hopwood, Orlando, & Clark, 2010).** The MPRD-DF was designed to detect malingered pain-related disability. Like the RDF, the MPRD-DF is also derived from a combination of discriminant function weighted scores from several of the PAI scales and indices. The MPRD-DF has not yet been validated across forensic or clinical populations; however, this scale is proposed to be valuable in an mTBI assessment context given the strong link found between pain and mTBI
cases (Beetar, Guilmette, & Sparadeo, 1996; Greve, Ord, Bianchini & Curtis, 2009; Nampiaparampil, 2008). Notably the NDS and MDRF-DF are research scales that are not included on the PAI score report (Morey, 1991) but may be useful in settings where over-reporting of symptoms that can persist after an mTBI is suspected.

Utility of PVTs and SVTs

As reviewed earlier, according to the National Academy of Neuropsychology (NAN) Policy and Planning Committee, without the use of PVTs and SVTs the results of a neuropsychological evaluation may be inaccurate, and the administration of PVTs and SVTs is seen as a necessary component of any neuropsychological evaluation (Bush, 2015). What remains to be examined empirically is which SVT scales and indices are best at differentiating false positives (i.e., those who over-report symptoms that can persist after mTBI) from true positives (i.e., those with bonafide symptoms that can persist after mTBI), and how these SVT scales and indices fair against PVTs in predicting over-reporting of symptoms that can persist after mTBI. To this end, a simulation study was conducted in which participants were administered a battery of PVTs and embedded SVTs under two different instructions sets: standard instructions or over-reporting instructions. The incremental increase in predictive validity of the PVTs and SVTs in differentiating those that completed these tests under standardized administration instructions versus those instructed to over-report symptoms that can persist after mTBI was examined.

The Current Study

The broad goals of this study were two-fold: (1) to determine which set of SVTs, from the MMPI-2-RF and PAI, are best in distinguishing those that over-report symptoms that may persist after mTBI versus those that respond “honestly”; and (2) to determine if SVTs could
provide significant incremental predictive validity beyond that of PVTs in the detection of over-reporting of these symptoms. An experimental analogue design was employed to address directly these goals. The independent variable was the instructional set provided to the participants standard (i.e., respond honestly) instruction [SI] versus. over-reporting instruction [OR]) and the dependent variables were the PVT and SVT scales and indices. The following hypotheses were tested through a series of hierarchical logistic regression models (see Table 5).

**Hypotheses**

**Comparative capacity of the MMPI-2-RF and PAI SVT scales and indices.** Hypotheses were based, in part, on a pilot study by Carpio, Zakzanis & Bagby, (unpublished) where participants were instructed to simulate symptoms that can persist after an mTBI on the PAI and MMPI-2-RF. Results from this pilot study revealed that the MMPI-2-RF over-reporting scales, in general, were better predictors of simulated over-reporting than the PAI over-reporting scales and indices. Based on these results, it is hypothesized that the MMPI-2-RF SVT scales will be better at differentiating SI from OR than the SVT scales and indices of the PAI.

It is also hypothesized that the RBS, FBS-r scales of the MMPI-2-RF and the NIM, RDS and MAL indices of the PAI will be the best predictors for differentiating SI from OR given the emphasis of these scales and indices on over-reporting of symptoms that can persist after an mTBI.

**Incremental Predictive Capacity of SVT scales and Indices to PVTs.** Based on the variety of symptoms that can persist after mTBI – cognitive, emotional, somatic – it is hypothesized that the SVT scales and indices of the MMPI-2-RF and PAI will add significant incremental validity in differentiating SI from OR.
Method

Participants

Sample Size. A statistical power analysis was performed using G*Power for sample size estimation (Faul et al., 2007). With an anticipated effect size of \( d = .50 \), alpha = .05 and power =0.80, the projected sample size needed was 135 (45 participants per group). This sample size estimates were also informed by previous studies employing similar strategies (see for e.g., Green, Kern, & Heaton, 2004) and by a meta-analysis of simulation studies using the MMPI-2 in student samples where sample sizes ranged from 22 to 179 (\( M_{\text{sample size}} = 52; \) Rogers, Sewell, & Salekin, 1994).

Participant Characteristics. Participants (76.6% women) were English-speaking and between the ages of 18 and 55 years (\( M_{\text{age}} = 18.96, SD = 2.53 \)) with no self-reported history of psychopathology or a neurological condition. Demographic information for the 164 valid participant protocols included in the current study (see Procedure section) is displayed in Table 1. All participants were recruited from the University of Toronto Scarborough online research participant pool who participated in the study in exchange for course credit.

Measures

Performance Validity Tests\(^1\)

Rey Fifteen Item Test (Rey-FIT; Rey, 1964). The Rey-FIT is a test to assess symptom validity of feigned memory impairment. Participants are asked to memorize and recall 15 items. The items are in five rows with three characters per line. Each item belongs to one of four categories and are presented as follows: A, B, C; 1, 2, 3; a, b, c; a circle, a square, a triangle; and

\(^1\) The three PVTs used in the initial development and validation of the RBS were not included in the current study to avoid conflation of the independent and dependent variables. The PVTs used in the development and validation of the RBS are the Test of Memory Malingering, the Word Memory Test, and the Computerized Assessment of Response Bias.
Roman numerals I, II, and III. After a short delay, participants are presented with a recognition task to identify the items they had seen previously. The potential range of raw scores on the Rey-FIT is 0-30. The recommended validity cutoff for the REY-FIT for suspected feigning is a raw score of \( \leq 18 \) (Lezak, 2004).

**Reliable Digit Span (RDS; (Greiffenstein, Baker, & Gola, 1994).** The RDS is a component of the Digit Span subtest of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) and the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987). The longest span of digits recalled without error (forward and backward) is recorded. The number of digits in this sequence is summed yielding a RDS score. The potential range of raw scores on the RDS is 0-16. The recommended validity cutoff for the RDS for suspected feigning is a raw score of \( \leq 7 \) (Mathias at al., 2002).

**Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997).** The VSVT is a 48 item, two-alternative forced choice computer administered recognition test, used to assess the validity of recounted cognitive complaints (e.g., memory problems). For each trial, examinees study a 5-digit number on a screen. After a short delay, examinees are represented with a two-alternative recognition trial in which they are asked to select the 5-digit number they had been presented with before. The potential range of “total correct” items is 0-48. The recommended cut-off for suspected feigning is a raw score \( \leq 30 \).

**Symptom Validity Scales and Indices**

**Minnesota Multiphasic Personality Inventory 2 (restructured form) (MMPI-2-RF; Ben-Porath & Tellegen, 2008).** The MMPI-2-RF is a 338-item self-report inventory, which assesses personality and psychopathology. The MMPI-2-RF is a multi-scale instrument assessing personality and psychopathology. Responses are indicated on a “forced choice” – true/false
format. The instrument is composed higher order, restructured clinical, specific problems, interest, personality psychopathology and validity. For the purposes of the current study, the validly scales that were designed to measure non-content based invalid responding – cannot say (CNS), variable response inconsistency (VRIN-r) and true response inconsistency (TRIN-r)– were used to eliminate invalid protocols. CNS, VRIN-r, TRIN-r were not used as predictors in regression analyses. Studies that indicate that those that who are deemed invalid based on VRIN-r and TRIN-r also tend to be invalid on other tests (Burchett & Bagby, 2014).

In addition, the validity scales designed to measure over-reporting - F-r, Fp-r, Fs, FBS-r, RBS - were included in the analyses. The range of raw scores for F-r is 0-32; for Fp-r the range is 0-27; for Fs the range is 0-16; for FBS-r the range is 0-30; for RBS the range is 0-28. The F-r, Fp-r, and Fs, do not overlap in content. FBS-r shares three items with Fs and one with Fp-r. The RBS and FBS-r share four items.

**Personality Assessment Inventory (PAI; Morey, 1991).** The PAI is a 344-item self-report inventory that assesses personality and psychopathology. Responses are indicated on a four-point Likert-scale (0 = False, 1 = Slightly True, 2 = Mainly True, 3 = Very True). The instrument is composed of clinical, interpersonal, treatment consideration and validity scales. For the purposes of the current study the scales that were designed to measure non-content based invalid responding – inconsistency (INC) and infrequency (INF) – were used to eliminate invalid protocols. In addition, the validity scales designed to measure over-reporting - NIM, MAL, RDF, NDS and MPRDF – were included in the analyses. The potential range of scores for NIM is 0-9. As described earlier, MAL index scores are computed from eight score configuration criteria that incorporate scores from several PAI scales and subscale; RDF and MPRDF were developed using discriminant function analyses procedures. As such, the range of the index
scores varies.

Post Experimental Questionnaire (Appendix A). A multiple-choice questionnaire was administered to assess the participants’ understanding of their task. Participants were asked to select from a possible four options that they believe best describes what their instructions were for the study. Participants who selected the option that reflects the instruction set that they were provided with were included in the sample. The second question addresses their level of confidence in being able to adhere to their instruction set. Participants who indicated that they were confident in their ability to adhere to instructions were included in the sample. This questionnaire served to identify those participants that were unable to identify what instructions they were given for completing the protocol. This questionnaire was explicitly used to determine the extent to which examinees understood and applied their instructions. Those that did indicated that they did not understand or were unable to apply the instructions that they were provided with were eliminated.

Procedure

A demographic form was administered to all participants to determine age, gender, number of years of education, ethnicity, and current/previous psychiatric and medical history. Participants were also asked whether English was their first language. If English was not their first language, participants were asked if they had completed and passed the Test of English as a Foreign Language (TOEFL; Educational Testing Services, 2017). The TOEFL measures the ability to use and understand written English at the university level. Participants were also screened for history of head injury. Questions informed by mTBI diagnostic criteria outlined by the mTBI committee of the head injury interdisciplinary special interest group of the American Congress of Rehabilitation Medicine were used ([ACRM]; 1993, e.g., loss of consciousness
(LOC), post traumatic amnesia (PTA), neuroimaging results, date since injury, diagnosis if applicable).

**Participant Flow and Elimination of Protocols.** A total of 216 undergraduate students signed up for the study (see Figure 1 below). Seven participants were eliminated due to technical issues (e.g., unable to complete part of the test, computer failure). Twenty-seven participants were eliminated because of non-content based invalid responding. Based on the MMPI-2-RF (Ben-Porath & Tellegen, 2008/2011) and PAI (Morey, 1991) administration and scoring manuals, protocols within the following criteria were excluded: CNS ≥ 15, VRIN-r or TRIN-r ≥ 80T and INC ≥ 73T and INF ≥ 75T. An additional 10 participants were excluded for selecting the incorrect instructional set or reported that they were not confident in their ability to follow their instructions on the post-experimental questionnaire (see Appendix A). An additional eight participants did not return for their second visit. A total of 52 participants were excluded, and 164 participants were retained (n = 83 in the standard instruction condition and n = 81 in the over-reporting instruction condition).

*Figure 1. Protocol Elimination.*
**Protocol administration.** Participants provided informed consent to participate in this two-part study upon arrival to the research laboratory at the University of Toronto. Participants who met inclusion criteria were randomly assigned to one of two conditions: a standard instructions condition or an over-reporting condition (see Figure 2 below). Participants in both conditions completed their respective tasks at two time points with an average of 7.4 days between Time 1 (T1) and Time 2 (T2). All participants completed a demographics form, an mTBI screening questionnaire and a post experimental questionnaire at both time points.

*Figure 2. Design and procedure.*

At T1, participants in the standard instruction condition completed the PVTs and the SVTs under standardized administration instruction. At T2, these participants completed the SVTs again under the same standardized administration instructions. At T1, participants in the over-reporting instruction condition completed the SVTs under standardized administration instructions. The PVTs were administered to each group only once as practice effects are common and could contaminate the results (Calamia, Markon & Tranel, 2012). At T2, these
participants completed the PVTs and SVTs as if they were suffering from symptoms can persist after an mTBI (i.e., cognitive impairment [Rohling et al, 2011], emotional problems [Mathias & Coats, 1999] and somatic complaints [Nampiapampil, 2008]).

These participants were instructed to simulate (“over-report”) these symptoms and were cautioned to avoid obvious dissimulation and detection. In order to inform participants on what symptoms to simulate, a description of mTBI and the symptoms that can persist following the injury was read to them. The participants were left with the description to refer to while they completed the protocol (see Appendix B). Participants were also provided with a case vignette describing a situation in which “faking” these symptoms might be beneficial and was relevant to students (i.e., faking symptoms to receive an extension for coursework; see Appendix C).

Statistical Analyses

Preliminary analyses. Paired sample t-tests were used to ensure that mean differences within groups were not due to the impact of retesting. The mean scores for the SVT scales and indices of the participants in the standard instruction condition at T1 and T2 were compared. These results indicated that the scores at T1 and T2 were “equivalent” on the SVT scales and indices (see Table 2). Significance was set at $p < 0.005$ (0.05/10) to control for family wise error. There were no significant differences between T1 and T2 in the standard instruction condition on all SVT scales and indices, resulting in minimal change on mean scores across the two time points. Retest correlations were also computed to determine the stability of scores at T1 and T2 in the standard instruction condition. Mean scale scores exhibited significant retest stability, with $r$s ranging from $.59 - .88$, $p < 0.001$.

Paired samples t-tests were used examine mean differences on the SVT scales and indices between two groups that completed the protocol under standard instructions at T1. This analysis
was conducted to ensure that group differences were not due to random assignment. These results indicated that the mean differences between these groups were not due to the impact of random assignment.

A multivariate analysis of variance (MANOVA) was performed to examine the mean differences between the two groups that completed the protocol under standard instructions at T1. The mean differences between participants in the standard instructions condition at T1 and the over-reporting condition at T2 on all PVT and SVT scales and indices were also examined to ensure that there were significant and meaningful differences on all PVT and SVT scales and indices. These differences would suggest that experimental manipulation was successful. The assumptions for MANOVA were tested and met beforehand.

**Comparative Predictive Capacity of PVTs and SVTs.** Participants in the standard instructions ‘honest’ condition at T1 (SI) and the over-reporting condition at T2 (OR) were included in the subsequent regression analyses. These two groups were used so that comparisons could be made for all of the PVT and SVT scales and indices, as these were the two groups that were administered all of these measures. The first set of regressions compared the predictive capacity of the two SVTs, while the second set examined the incremental increase in predictive capacity of the SVTs among the PVTs.

**Results**

**Group Differences at Time 1**

The means, standard deviations, and estimates of effect size (e.g., Cohen’s $d$) for each SVT scale and index of each group at T1 are displayed in Table 3. The differences between these groups on all scales and indices at T1 were not significant, $F(10, 153) = 1.06, p = 0.40$; Pillai’s Trace = 0.065, partial $\eta^2 = .07$), establishing relative equivalence between these baseline
measurements. Effect sizes of 0.20, 0.50 and 0.80 represent small, medium and large effect sizes respectively (Cohen, 1988). Cohen’s $d$ ranged between 0.01 and 0.20 ($m = 0.12$), reflecting a negligible difference.

**Difference Between Experimental Groups Based on Instructional Set**

The means, standard deviations, and estimates of effect size of the PVT and SVT scale and index scores of the SI and OR groups are displayed in Table 4. The differences between the SI and OR groups on all PVT and SVT scales and indices were statistically significant, $F (13, 168) = 33.33, p < .0001; \text{Pillai's Trace} = 0.743$, partial $\eta^2 = .74)$. Estimates of effect size ranged between 0.89 and 2.77 with a mean effect size estimate of 1.74. Effect size estimates in this range represent large, meaningful differences (see Zakzanis, 1998 for a review).

**Hierarchical Binary Logistic Regression Analyses**

The assumptions of binary logistic regression were tested and met. All summary statistics for the series of regression analyses are displayed in Table 5. Effect size estimates for $R^2$ change are based on the guidelines provided by Hunsley and Meyer (2003) which indicate that .10, .30, and .50 are small, medium and large effects, respectively. The values of overall classification correct (OCC), sensitivity (SN), specificity (SP), positive predicative value (PPV) and negative predictive value (NPV) for the models tested (see regression model summary section below) are also displayed in Table 5.²

**Summary of Models Tested.** To examine the predicative capacity of the PVTs and SVT scales and indices, nine of hierarchical logistic regression models were tested. In Table 5 the nine models tested in three sets of regression analyses are displayed. In the first set of regression

---

² Sensitivity refers to the proportion of individuals correctly classified as over-reporting (e.g., true positive rate), whereas specificity refers to the proportion of individuals correctly classified as not over-reporting (e.g., true negative rate). Positive predictive value represents the probability that a participant is over-reporting given a positive test result whereas represents the probability that a person is not over-reporting given a negative test results.
models (i.e., Models 1, 2 and 3), the focus was on testing the capacity of the MMPI-2-RF SVT scales predictive capacity to distinguish the over-reporting from honest profiles, on its own, and relative to the PAI SVTs and the PVTs. In Model 1 only the MMPI-2-RF SVTs were entered as a block to determine the predictive capacity of these SVTs only. In Model 2, a second block was added in which PAI SVTs are entered as a block after the MMPI-2-RF SVTs; this model was designed to test if the PAI SVTs add significant predictive capacity to the MMPI-2 SVTs. In the final model of this set — Model 3, a second block was added in which the PVTs are entered as a block after the MMPI-2-RF SVTs; this model was designed to test if the PVTs add any significant predictive capacity to the MMPI-2-RF SVTs.

In the second set of regression models (i.e., Models 4, 5 and 6) the focus was on testing the capacity of the PAI SVT scales and indices predictive capacity to distinguish the over-reporting from honest profiles, on its own, and relative to the MMPI-2-RF SVT scales and the PVTs. In Model 4 only the PAI SVTs were entered as a block to determine the predictive capacity of these SVTs. In Model 5, a second block was added in which the MMPI-2-RF SVTs are entered as a block after the PAI SVTs; this model was designed to test if the MMPI-2-RF SVTs add significant predictive capacity to the PAI SVTs. In the final model of this set, Model 6, a second block was added in which the PVTs were entered as a block after the PAI SVTs; this model was designed to test if the PVTs add any significant predictive capacity to the PAI SVTs.

In the third set of regression models (i.e., Models 7, 8 and 9) the focus was on testing the incremental predicative capacity of the MMPI-2-RF and PAI SVT scales and indices to distinguish the over-reporting from honest profiles in comparison to the PVTs. In Model 7 only the PVTs were entered as a block to determine the predictive capacity of these tests. In Model 8, a second block was added in which the MMPI-2-RF SVTs were entered as a block after the
PVTS; this model was designed to test if the MMPI-2-RF SVTs add significant predictive capacity to the PVTs. In the final model of this set, Model 9, a second block was added in which the PAI SVTs were entered as a block after the PVTs; this model was designed to test if the PAI SVTs added any significant predictive capacity to the PVTs. Scales that remained statistically significant in the regression equations would indicate incremental predictive validity. Diagnostic efficiency statistics were also computed for each of the nine models to assess SN, SP, NPV, PPV, and OCC.

**Comparative Predictive Capacity of the MMPI-2-RF and the PAI SVT Scales and Indices.** The SVT scales of the MMPI-2-RF were entered in the first block of the regression and this model was significant ($\chi^2 (5, N = 164) = 150.17, p < 0.001$). When the SVT scales and indices of the PAI were entered into the second block, the change in model fit was not significant ($\chi^2 (3, N = 164) = 7.51, p = 0.19$), indicating that the PAI SVT scales indices do not add significant incremental validity in differentiating SI from OR above the MMP-2-RF SVT scales. Overall, SN, SP, NPV and PPV statistical values are strong when only the MMPI-2-RF SVT scales were entered into the model (see Table 5). The Fs and RBS scales of the MMPI-2-RF were identified as significant predictors in this model confirming the a priori hypothesis. Fs and RBS were associated with increased likelihood of being correctly classified as OR. The summary statistics for each predictor are displayed in Table 6.

The block entry of predictors was reversed to examine whether the MMPI-2-RF SVT scales added incremental validity over the PAI scales and indices in differentiating SI from OR. The PAI SVT scales and indices were entered first, and this model was significant ($\chi^2 (5, N = 164) = 127.18, p < 0.001$). SN, SP, NPV and PPV values are strong, but are inferior to the diagnostic efficiency statistics of the model that only included the SVT scales of the MMPI-2-RF (e.g.,
Model 1). The MMPI-2-RF SVT scales were entered in the second block and the change in model fit was significant ($\Delta \chi^2 = 30.50, p < .001$), indicating that the MMPI-2-RF SVT scales add significant incremental validity above the PAI in differentiating SI from OR. This incremental increase is associated with a medium effect size (e.g., change in $R^2$; see Table 5). The NIM scale of the PAI was a significant predictor in this model. Odds ratios suggest that an increase in NIM is associated with increased likelihood of being correctly classified as OR. When the OR scales of the MMPI-2-RF were added, the classification accuracy was comparable to Model 1 (see Table 5). Specifically, the RBS scale was identified as the single significant predictor among all SVT scales and indices. Odds ratios suggest that an increase in RBS is associated with increased likelihood of being correctly classified as OR. The summary statistics for each predictor are displayed in Table 6.

**Incremental Predictive Capacity of the MMPI-2-RF SVT Scales and the PVTs.** The PVTs were entered in the first block and the model fit was significant ($\chi^2 (3, N = 164) = 125.28, p < 0.001$). SN, SP, NPV and PPV values were strong but are inferior to Model 1, which included only the SVT scales on the MMPI-2-RF. The MMPI-2-RF SVT scales were entered in the second block, and the change in model fit was significant ($\Delta \chi^2 = 50.27, p < .001$) indicating that the MMPI-2-RF SVT scales add significant incremental validity above the PVTs in differentiating SI from OR. This incremental increase in predictive capacity was associated with a large effect size (see Table 5) and the classification accuracy was superior to all of the models that were tested. Specifically, the RDS and VSVT were identified as significant predictors among the PVTs, and the RBS scale of the MMPI-2-RF was identified as the significant predictor among the SVT scales. Odds ratios indicate that lower scores on the VSVT and RDS were associated with increased likelihood of being correctly classified as OR whereas higher
scores on the RBS scale increased the likelihood of being correctly classified as OR. These results demonstrate that the PVTs outperform the PAI SVT scales and indices, but not the MMPI-2-RF SVT scales with respect to diagnostic efficiency. The summary statistics for each predictor are displayed in Table 7.

The block entry of predictors was reversed to examine whether the PVTs added incremental validity over the MMPI-2-RF SVT scales in differentiating SI from OR. The MMPI-2-RF SVT scales entered in first block have been discussed previously (see Model 1, Table 5). The PVTs were entered in the second block and the change in model fit was significant, ($\Delta\chi^2 = 25.40, p < .001$) indicating that the PVTs add significant incremental validity to the MMPI-2-RF SVT scales in differentiating SI from OR. The incremental increase was associated with a medium effect size (see Table 5). Overall, SN, SP, NPV and PPV values are strong, and no additional significant predictors emerged after reverse entry - the RDS and VSVT were identified as significant predictors among the PVTs, and the RBS scale of the MMPI-2-RF was identified as a significant predictor among the SVT scales. Odds ratios indicated that lower scores on the VSVT and RDS were associated with increased likelihood of being correctly classified as OR whereas higher scores on the RBS scale increased the likelihood of being correctly classified as OR. The summary statistics for each predictor are displayed in Table 7.

**Predictive Capacity of the PAI SVT Scales and Indices and PVTs**

In order to provide a full picture of the comparative predictive capacity of SVT indices with PVTs, the SVT indices of the PAI were also examined. In some cases, clinicians may opt to use the PAI over the MMPI-2-RF based on clinical preference regardless of the outperformance of the MMPI-2-RF SVT scales.

The PVTs entered into the regression model in the first block have been discussed previously
(see Model 7, Table 5). The PAI SVT scales and indices were entered in the second block and the change in model fit was significant ($\Delta \chi^2 = 39.21$, $p < .001$), indicating that the PAI SVT scales and indices add significant incremental validity in differentiating SI from OR. The incremental increase was associated with a medium effect size (see Table 5). Overall, SN, SP, NPV and PPV values are strong. Specifically, the RDS and VSVT were identified as significant predictors among the PVTs, and the NIM scale of the PAI was identified as a significant predictor among the SVT indices. Odds ratios indicated that higher scores on the VSVT and RDS were associated with decreased likelihood of being correctly classified as an OR whereas higher scores on the NIM scale increased the likelihood of being correctly classified as OR. The summary statistics for each predictor are displayed in Table 8.

The block entry of predictors was reversed to examine whether the PVTs added incremental validity over the PAI SVT scales and indices in differentiating SI from OR. The PVTs were entered into the second block and the change in model fit was significant, ($\Delta \chi^2 = 37.32$, $p < .001$) indicating that the PVTs add significant incremental validity in differentiating SI and OR. The incremental increase by adding the PVTs was associated with a medium effect size. Overall, SN, SP, NPV and PPV values are strong. Specifically, the RDS and VSVT were identified as significant predictors among the PVTs, and the NIM scale of the PAI remained as the single significant predictor among the SVT indices. Odds ratios indicated that lower scores on the VSVT and RDS were associated with decreased likelihood of being correctly classified as an OR whereas higher scores on the NIM scale increased the likelihood of being correctly classified as OR. The summary statistics for each predictor are displayed in Table 8.

**Discussion**

Psychologists have available to them various psychometric tests to assess the breadth,
severity and veracity of symptoms. Ideally, test choice should be informed empirical evidence (Hunsley & Mash, 2007), yet it remains unclear what set of tests are best to assess the over-reporting of symptoms that can persist after an mTBI. No published study has yet examined which of the various SVT scales and indices on both the MMPI-2-RF and PAI are the best predictors for the detection of over-reporting of these symptoms or examined the incremental validity of SVT scales and indices of the MMPI-2-RF and PAI above that of the PVTs in the context of mTBI. In the current study, an experimental simulation design was employed to evaluate the extent that SVTs add incremental validity to PVTs in predicting those that are administered tests under standard instructions and those instructed to over-report symptoms that can persist following an mTBI.

The MMPI-2-RF SVT scales were better predictors of group membership than the PAI SVT scales and indices. Specifically, two of the scales - Fs and RBS – were significant predictors regardless of the order of entry of variables. An increase in scores on either of these scales increases the probability of being correctly identified as part of the over-reporting group. Although there were significant predictors among the PAI SVT scales and indices, an increase in these scales resulted in decreased classification accuracy. When examining the incremental validity of the MMPI-2-RF SVT scales in comparison to the PVTs, results indicate that the MMPI-2-RF SVT scales add predictive validity beyond that of the PVTs. In addition, 93% of individuals were correctly predicted as over-reporting and 96% were correctly predicted as not over-reporting, suggesting that including the MMPI-2-RF in addition to PVTs would be useful in differentiating between false positive and true positives with respect to over-reporting symptoms in the context of a multi-method assessment. In this study, the probability of being identified as an “over-reporter” increased with each unit increase in the RBS scale and with each unit
decrease in the VSVT and RDS. These results are consistent with the literature, where higher scores on RBS are indicative of possible feigning of cognitive symptoms (Ben-Porath, 2012), and lower scores on the VSVT and RDS are indicative of possible feigning and/or suboptimal effort (Slick et al., 2004). In addition, this combination of tests provided the highest levels of specificity and PPV with significant increases in model fit statistics.

Although the first set of regression analyses (i.e., Models 1, 2, and 3) demonstrated that the MMPI-2-RF SVT scales provided more incremental predictive utility than the PAI SVT scales and indices, some clinicians may still opt to use the PAI due to simple clinical preference; rather than using the empirically validated scales of the MMPI-2-RF SVT scales as shown by the results of the current study. For those who choose to use the PAI, the NIM scale of the PAI was identified as the single best predictor among all PAI SVTs and PVTs, although the RDS and VSVT added significant predictive capacity to the NIM scale.

To help clarify why the RBS was the best predictor in this study, even when compared to PVTs, it is important to examine how this scale was developed. Bootstrapping comparisons were used to develop the scale, which improve diagnostic efficiency statistics (Rogers & Bender, 2003). Multiple measurements (i.e., three well validated PVTs) of over-reporting were used to classify groups in the development of this scale and therefore, maximized external validity. The RBS was specifically developed using SVT performance from a forensic disability sample as the referent criterion, and employed multiple regression analyses to select a set of items that would predict failure on three PVTs (Gervais et al., 2007). Thus, the development of the RBS makes it probable that this scale would outperform the PVTs used in the current study. The other SVT scales and indices were not developed in this manner, but rather employed strategies such as infrequent item endorsement, discriminant function analysis, rational item selection, and multiple
regression analysis. In addition, it may be the case that cognitive complaints are the “cardinal” symptoms that may persist after an mTBI (e.g., memory, concentration). As such, a scale that was designed to measure over-reporting cognitive impairment would likely outperform scales that assess general over-reporting, or over-reporting of psychiatric symptoms in the context of mTBI.

In addition, the PVTs used in the current study measure over-reporting of memory complaints, rather than the wide range of cognitive symptoms that may persist after an mTBI. Therefore, the RBS may have yet another advantage as the content that this scale covers is not exclusive memory complaints, as is the case with the PVTs. Future studies could utilize PVTs that measure over-reporting of other cognitive symptoms and compare the RBS to these other tests.

The diagnostic efficiency statistics for the model that included both the PVT and MMPI-2-RF SVT scales in this study are strong, yet it is important to note that these statistics are influenced by base rates. In contrast to sensitivity and specificity, PPV and NPV are affected by the actual base rate of the condition in question. If base rates are higher in certain contexts (e.g., clinical settings) it is more likely that an examinee will be correctly identified as over-reporting. On the other hand, if the base rates are lower it is less likely that an examinee will be correctly identified as over-reporting. The base rate of over-reporting in this study was derived from the over-reporting group in the current study, and was approximately 50% (49.39% based on the experimental randomization). This base rate may be higher than the actual base rates of over-reporting of symptoms that can persist after an mTBI; the base rate for over-reporting mTBI related symptoms reported in the literature range between 10% and 40% (Mittenberg et al., 2002), which suggests that the PPV and NPV reported in the current study may be lower in other
Administering the RBS, VSVT and RDS in an assessment may reduce the probability of obtaining a false positive, which is, incorrectly identifying someone as an over-reporter when they are not. Of particular importance in the context of an assessment is reducing false positives with respect to over-reporting. The consequences of falsely classifying an examinee as over-reporting do outweigh the consequences of “missing” the response style. Such interpretation can therefore impact the decisions that are made from psychological assessments that include disqualifying an examinee from needed compensation, disability benefits, or treatment.

The results of this analogue investigation suggest that including the RBS SVT scale of the MMPI-2-RF and two of the three PVTs as part of an assessment for over-reporting of symptoms following an mTBI would be clinically useful, although replication in a clinical sample is needed to substantiate this. Replication in a clinical sample would assess the degree to which the results of the current study are valid and reliable. In addition, it would examine whether or not these results apply to bonafide clinical sample. Yet the results demonstrate that there are clear differences in the predictive capacity of two commonly used self-report inventories and that the MMPI-2-RF SVT scales outperform those of the PAI. The results also demonstrate that the MMPI-2-RF SVT scales outperform the PVTs. Based on the results of this study, it is recommended that the MMPI-2-RF be administered to examinees who are suspected of over-reporting of symptoms following an mTBI. Administering PVTs following the MMPI-2-RF may be useful to substantiate findings based on the MMPI-2-RF SVT indices; based on the results of this study, additional tests are not necessary to improve predictive capacity. Yet, the value of incremental validity through multimethod assessment has been formulated by Campbell

---

3 Some research has illustrated that base rates approach 80% although this is likely due to an artifact based on loose diagnostic criteria (Carroll et al., 2004; Mittenberg et al., 2010; Ord et al., 2010).
and Fiske in 1959; and a clinician should not rely on a single metric (e.g., one test, one scale, or one index) to arrive at a diagnostic conclusion (Hopwood and Bornstein, 2014). Specific patterns that increase confidence in the validity of an assessment include consistency between self-report, performance, medical history, and informant report (IOM, 2015).

Although this study has several important implications with respect to instrument selection in the assessment of mTBI; there are some are limitations. Simulation research, which is deeply steeped in the experimental method, has strong internal validity; this method includes standardized instructions, different experimental conditions, incentives and manipulation checks, all of which minimize error (Rogers, 2008; Rogers & Cruise, 1998), but potentially compromise the generalizability of the results to applied contexts. Although experimental control and standardization of procedures are methodologically superior to other designs with less error reducing control (e.g., ‘known groups’ design), explicitly instructing participants to over-report is not facsimile to clinically motivated over-reporting. Future investigations may consider using known-groups design that more closely mimic clinical contexts.

There are also limitations of the utility of self-report inventories with embedded SVT indices in some populations. The MMPI-2-RF and PAI require minimum reading and comprehension levels (i.e., grades eight and four respectively) restricting the use of these instruments in populations with lower reading and comprehension abilities. Although there are audio versions of these tests, audio versions would not account for word comprehension issues. Here instead, PVTs may be useful for these populations where reading is not required to complete the tests.

In addition, the undergraduate sample used in this study may limit the generalizability of the results as it pertains to other age groups (e.g., older adults). Future investigations should use a
sample that includes more diverse age groups, or use other distinct age groups when aiming to replicate these results; however, it is important to note that young adults are disproportionately represented in incidents of mTBI compared to most other age groups (Centers for Disease Control and Prevention [CDC], 2003); the population that was examined in this study is fairly consistent with the age prevalence rates of mTBI.

Although this study addresses a gap in the research literature with respect to instrument selection, it is important to note that SVT and PVTs only alert clinicians to the possibility of over-reporting. These scales and indices are not used to conclusively indicate the presence or absence of it. No single measure or index should be used to determine “feigned” symptoms (Marshall et al., 2010; Schroeder & Marshall, 2010). In a neuropsychological evaluation, including multiple scales indices that measure over-reporting to substantiate other test findings are imperative. In order to achieve maximal levels of confidence in test data and the interpretations dawn from a psychological assessment over-reporting response style should be assessed objectively, while supplementing with clinical history, interview and informant report. Overall, based on the collective results of the current study, it is recommended that the MMPI-2-RF be administered as part of a neuropsychological evaluation, and that the RBS scores are used as an SVT when assessing for over-reporting of symptoms that may persist following an mTBI.
References


Rehabilitation Medicine, 36(0), 28-60.


Franzen, M. D., Iverson, G. L., & McCracken, L. M. (1990). The detection of malingering in


Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and
Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology, 18*(2), 86-94.


Rohling, M. L., Larrabee, G. J., Greiffenstein, M. F., Ben-Porath, Y. S., Lees-Haley, P., Green,


van Gorp, W. G., Humphrey L.A., Kalechstein, A.L., Brumm, V.L., McMullen, W.J., Stoddard,


Table 1.

*Participant self-reported ethnicity.*

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>Number of Valid Protocol* Participants</th>
<th>Percentage (n = 164)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White/Caucasian</td>
<td>19</td>
<td>11.6</td>
</tr>
<tr>
<td>East Asian</td>
<td>32</td>
<td>19.5</td>
</tr>
<tr>
<td>South Asian</td>
<td>69</td>
<td>42.1</td>
</tr>
<tr>
<td>Black or of African or Caribbean</td>
<td>14</td>
<td>8.5</td>
</tr>
<tr>
<td>Latin American</td>
<td>2</td>
<td>1.2</td>
</tr>
<tr>
<td>Middle Eastern</td>
<td>15</td>
<td>9.1</td>
</tr>
<tr>
<td>Mixed Race/Ethnicity</td>
<td>13</td>
<td>7.9</td>
</tr>
<tr>
<td>Preferred not to disclose</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

*Note.* *valid participant protocols included in the current study are based on the inclusion criteria and elimination of protocols outlined in the procedure section. These values are rounded to the nearest whole number.
Table 2.
Means, standard deviations, and within group comparisons of standard instruction condition group at time 1 and time 2.

<table>
<thead>
<tr>
<th>Validity Index</th>
<th>Standard Instruction Condition T1 (n = 83)</th>
<th>Standard Instruction Condition T2 (n = 83)</th>
<th>Within-subject comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>r</td>
</tr>
<tr>
<td>MMPI-2-RF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-r</td>
<td>4.60 (4.24)</td>
<td>4.29 (4.18)</td>
<td>0.88</td>
</tr>
<tr>
<td>Fp-r</td>
<td>2.47 (1.68)</td>
<td>2.47 (1.86)</td>
<td>0.82</td>
</tr>
<tr>
<td>Fs</td>
<td>2.13 (1.87)</td>
<td>1.89 (2.02)</td>
<td>0.66</td>
</tr>
<tr>
<td>FBS</td>
<td>9.53 (3.81)</td>
<td>9.17 (3.96)</td>
<td>0.84</td>
</tr>
<tr>
<td>RBS</td>
<td>7.31 (3.61)</td>
<td>7.31 (3.33)</td>
<td>0.81</td>
</tr>
<tr>
<td>PAI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM</td>
<td>2.82 (2.74)</td>
<td>2.46 (2.48)</td>
<td>0.82</td>
</tr>
<tr>
<td>MALI</td>
<td>1.02 (0.94)</td>
<td>1.01 (0.88)</td>
<td>0.59</td>
</tr>
<tr>
<td>RDF</td>
<td>-0.56 (1.07)</td>
<td>-0.48 (1.02)</td>
<td>0.66</td>
</tr>
<tr>
<td>NDS</td>
<td>3.16 (1.95)</td>
<td>3.33 (2.41)</td>
<td>0.69</td>
</tr>
<tr>
<td>MPRDF</td>
<td>20.73 (3.51)</td>
<td>20.62 (3.86)</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Note: Mean and Standard deviations of raw scores for each scale and index are presented. F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias Scale; NIM = Negative Impression Management; MAL = Malingering Index; RDS = Roger’s Discriminant Function; NDS = Negative Distortion Scale; MPRDF = Malingered Pain-Related Disability Function. M = mean; SD = standard deviation; r = retest correlations between time 1 and time 2; d = effect size. *p < .05; **p < .01; ***p < .001.
### Table 3.

**Means, standard deviations, F tests, and effect size estimates between groups at time 1**

<table>
<thead>
<tr>
<th>Validity Index</th>
<th>Standard Instruction Condition T1</th>
<th>Over-reporting Condition T1</th>
<th>Between subject comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>F (10, 153) d 95% CI</td>
</tr>
<tr>
<td><strong>MMPI-2-RF</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-r</td>
<td>4.70 (4.44)</td>
<td>3.87 (3.92)</td>
<td>0.63</td>
</tr>
<tr>
<td>Fp-r</td>
<td>2.53 (1.81)</td>
<td>2.03 (1.63)</td>
<td>2.37</td>
</tr>
<tr>
<td>Fs</td>
<td>2.22 (2.02)</td>
<td>2.02 (1.86)</td>
<td>0.38</td>
</tr>
<tr>
<td>FBS-r</td>
<td>9.60 (3.88)</td>
<td>9.68 (4.15)</td>
<td>0.77</td>
</tr>
<tr>
<td>RBS</td>
<td>7.25 (3.58)</td>
<td>7.41 (3.73)</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>PAI SVT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM</td>
<td>3.10 (3.30)</td>
<td>2.61 (3.03)</td>
<td>0.37</td>
</tr>
<tr>
<td>MALI</td>
<td>1.05 (0.99)</td>
<td>0.97 (0.90)</td>
<td>0.91</td>
</tr>
<tr>
<td>RDF</td>
<td>-0.06 (1.07)</td>
<td>-0.06 (0.96)</td>
<td>0.26</td>
</tr>
<tr>
<td>NDS</td>
<td>3.25 (2.17)</td>
<td>3.22 (1.98)</td>
<td>0.00</td>
</tr>
<tr>
<td>MPRDF</td>
<td>20.51 (3.38)</td>
<td>21.30 (3.38)</td>
<td>0.13</td>
</tr>
</tbody>
</table>

\[
M = 0.12
\]

*Note:* Mean and Standard deviations of raw scores for each scale and index are presented. REY FIT = Rey Fifteen Item Test; RDS = Reliable Digit Span; VSVT = Victoria Symptom Validity Test; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias Scale; NIM = Negative Impression Management; MAL = Malingering Index; RDS = Roger’s Discriminant Function; NDS = Negative Distortion Scale; MPRDF = Malingered Pain-Related Disability Function. \( M = \) mean; \( SD = \) standard deviation; \( d = \) Cohen’s d. *\( p < .05; **p < .01; ***p < .001. \]
Table 4.
Means, standard deviations, and between group comparisons of standard instruction (SI) and over-reporting instruction (OR) groups.

<table>
<thead>
<tr>
<th>Validity Index</th>
<th>Standard Instruction</th>
<th>Over-reporting Instruction</th>
<th>Between subject comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 (SI)</td>
<td>T2 (OR)</td>
<td>F (13, 150)</td>
</tr>
<tr>
<td>PVT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REY FIT</td>
<td>14.55 (1.14)</td>
<td>12.73 (2.41)</td>
<td>38.67***</td>
</tr>
<tr>
<td>RDS</td>
<td>10.22 (1.98)</td>
<td>5.75 (3.23)</td>
<td>114.75***</td>
</tr>
<tr>
<td>VSVT</td>
<td>46.01 (6.31)</td>
<td>29.86 (10.84)</td>
<td>136.75***</td>
</tr>
<tr>
<td>MMPI-2-RF SVT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-r</td>
<td>4.29 (4.18)</td>
<td>18.23 (8.03)</td>
<td>195.70***</td>
</tr>
<tr>
<td>Fs</td>
<td>1.89 (2.02)</td>
<td>9.95 (3.95)</td>
<td>273.02***</td>
</tr>
<tr>
<td>Fs</td>
<td>1.89 (2.02)</td>
<td>9.95 (3.95)</td>
<td>273.02***</td>
</tr>
<tr>
<td>FBS</td>
<td>9.17 (3.96)</td>
<td>19.16 (4.62)</td>
<td>221.11***</td>
</tr>
<tr>
<td>RBS</td>
<td>7.31 (3.33)</td>
<td>18.68 (4.74)</td>
<td>316.86***</td>
</tr>
<tr>
<td>PAI SVT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM</td>
<td>2.46 (2.48)</td>
<td>13.63 (6.80)</td>
<td>196.90***</td>
</tr>
<tr>
<td>MALI</td>
<td>1.01 (0.88)</td>
<td>2.79 (1.70)</td>
<td>71.80***</td>
</tr>
<tr>
<td>RDF</td>
<td>-.48 (1.02)</td>
<td>0.56 (1.30)</td>
<td>32.23***</td>
</tr>
<tr>
<td>NDS</td>
<td>2.33 (2.41)</td>
<td>8.98 (5.45)</td>
<td>74.38***</td>
</tr>
<tr>
<td>MPRDF</td>
<td>20.62 (3.86)</td>
<td>26.52 (6.95)</td>
<td>45.55***</td>
</tr>
</tbody>
</table>

Note: Mean and Standard deviations of raw scores for each scale and index are presented. REY FIT = Rey Fifteen Item Test; RDS = Reliable Digit Span; VSVT = Victoria Symptom Validity Test; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias Scale; NIM = Negative Impression Management; MAL = Malingering Index; RDS = Roger’s Discriminant Function; NDS = Negative Distortion Scale; MPRDF = Malingered Pain-Related Disability Function. M = mean; SD = standard deviation; d = Cohen’s d. *p < .05; **p < .01; ***p < .001.
### Table 5.
Model characteristics and diagnostic efficiency statistics.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Model</th>
<th>χ²</th>
<th>Δ χ²</th>
<th>R²</th>
<th>Δ R²</th>
<th>R</th>
<th>OCC</th>
<th>SN</th>
<th>95% CI</th>
<th>SP</th>
<th>95% CI</th>
<th>PPV</th>
<th>95% CI</th>
<th>NPV</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI vs. OR Group (BR = 49.39)</td>
<td>1. MMPI-2-RF</td>
<td>150.17***</td>
<td>0.80</td>
<td>0.93</td>
<td>0.94</td>
<td>0.83 – 0.96</td>
<td>0.94</td>
<td>0.87 – 0.98</td>
<td>0.94</td>
<td>0.86 – 0.97</td>
<td>0.92</td>
<td>0.85 – 0.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. MMPI-2-RF + PAI</td>
<td>157.68***</td>
<td>7.51</td>
<td>0.82</td>
<td>0.02</td>
<td>0.14</td>
<td>0.93</td>
<td>0.96</td>
<td>0.83 – 0.96</td>
<td>0.96</td>
<td>0.90 – 0.99</td>
<td>0.96</td>
<td>0.89 – 0.99</td>
<td>0.92</td>
<td>0.85 – 0.96</td>
<td></td>
</tr>
<tr>
<td>3. MMPI-RF + PVT</td>
<td>175.56***</td>
<td>25.40***</td>
<td>0.88</td>
<td>0.08</td>
<td>0.28</td>
<td>0.95</td>
<td>0.93</td>
<td>0.85 – 0.97</td>
<td>0.96</td>
<td>0.90 – 0.99</td>
<td>0.96</td>
<td>0.89 – 0.99</td>
<td>0.93</td>
<td>0.86 – 0.97</td>
<td></td>
</tr>
<tr>
<td>4. PAI</td>
<td>127.18***</td>
<td>0.72</td>
<td>0.87</td>
<td>0.80</td>
<td>0.70 – 0.88</td>
<td>0.93</td>
<td>0.85 – 0.97</td>
<td>0.92</td>
<td>0.83 – 0.96</td>
<td>0.93</td>
<td>0.75 – 0.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. PAI + MMPI-2-RF</td>
<td>157.68***</td>
<td>30.50***</td>
<td>0.82</td>
<td>0.10</td>
<td><strong>0.32</strong></td>
<td>0.93</td>
<td>0.91</td>
<td>0.83 – 0.96</td>
<td>0.94</td>
<td>0.87 – 0.98</td>
<td>0.94</td>
<td>0.86 – 0.97</td>
<td>0.92</td>
<td>0.85 – 0.96</td>
<td></td>
</tr>
<tr>
<td>6. PAI + PVT</td>
<td>164.50***</td>
<td>37.32***</td>
<td>0.84</td>
<td>0.12</td>
<td><strong>0.35</strong></td>
<td>0.95</td>
<td>0.93</td>
<td>0.84 – 0.97</td>
<td>0.96</td>
<td>0.90 – 0.99</td>
<td>0.96</td>
<td>0.89 – 0.99</td>
<td>0.93</td>
<td>0.85 – 0.97</td>
<td></td>
</tr>
<tr>
<td>7. PVT</td>
<td>125.29***</td>
<td>0.71</td>
<td>0.90</td>
<td>0.86</td>
<td>0.77 – 0.93</td>
<td>0.94</td>
<td>0.87 – 0.98</td>
<td>0.93</td>
<td>0.86 – 0.97</td>
<td>0.88</td>
<td>0.80 – 0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. PVT + MMPI-2-RF</td>
<td>175.56***</td>
<td>50.27***</td>
<td>0.88</td>
<td>0.17</td>
<td><strong>0.41</strong></td>
<td>0.95</td>
<td>0.93</td>
<td>0.84 – 0.97</td>
<td>0.96</td>
<td>0.90 – 0.99</td>
<td>0.96</td>
<td>0.89 – 0.99</td>
<td>0.93</td>
<td>0.86 – 0.97</td>
<td></td>
</tr>
<tr>
<td>9. PVT+PAI</td>
<td>164.50***</td>
<td>39.21***</td>
<td>0.84</td>
<td>0.13</td>
<td><strong>0.36</strong></td>
<td>0.95</td>
<td>0.93</td>
<td>0.84 – 0.97</td>
<td>0.96</td>
<td>0.90 – 0.99</td>
<td>0.96</td>
<td>0.89 – 0.99</td>
<td>0.93</td>
<td>0.86 – 0.97</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Nagelkerke R² estimation was used for logistic regression. 
BR base rate, SN sensitivity, SP specificity, OCC overall correct classification, PPV positive predictive value, NPV negative predictive value; CI = confidence interval. Bold values indicate medium to large effect size estimates. SI = standard instructions Time 1; OR = over-reporting instructions Time 2. *p<0.05, **p<0.01, ***p<0.001.
Table 6. Hierarchical Logistic Regression Analysis comparing the MMPI-2-RF and PAI scales and indices differentiating SI from OR groups.

<table>
<thead>
<tr>
<th>Validity Index</th>
<th>MMPI-2-RF</th>
<th>Block 1</th>
<th></th>
<th>Block 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>B</td>
<td>Odds Ratio</td>
<td>95% C.I.</td>
</tr>
<tr>
<td>F-r</td>
<td>-0.15</td>
<td>0.10</td>
<td>0.86</td>
<td>0.70</td>
<td>- 1.05</td>
</tr>
<tr>
<td>Fp-r</td>
<td>0.13</td>
<td>0.15</td>
<td>1.14</td>
<td>0.85</td>
<td>- 1.53</td>
</tr>
<tr>
<td>F-s</td>
<td>0.37*</td>
<td>0.17</td>
<td>1.45</td>
<td>1.05</td>
<td>- 2.00</td>
</tr>
<tr>
<td>FBS</td>
<td>-0.04</td>
<td>0.10</td>
<td>0.96</td>
<td>0.79</td>
<td>- 1.17</td>
</tr>
<tr>
<td>RBS</td>
<td>0.41**</td>
<td>0.14</td>
<td>1.50</td>
<td>1.15</td>
<td>- 1.97</td>
</tr>
<tr>
<td>PAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALI</td>
<td>-0.30</td>
<td>0.40</td>
<td>0.74</td>
<td>0.34</td>
<td>- 1.62</td>
</tr>
<tr>
<td>RDF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDS</td>
<td>-0.24</td>
<td>0.16</td>
<td>0.79</td>
<td>0.58</td>
<td>- 1.08</td>
</tr>
<tr>
<td>MPRDF</td>
<td>0.10</td>
<td>0.08</td>
<td>1.10</td>
<td>0.94</td>
<td>- 1.30</td>
</tr>
<tr>
<td>Reverse Entry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIM</td>
<td>0.49***</td>
<td>0.10</td>
<td>1.63</td>
<td>1.36</td>
<td>- 1.97</td>
</tr>
<tr>
<td>MALI</td>
<td>-0.29</td>
<td>0.31</td>
<td>0.75</td>
<td>0.41</td>
<td>- 1.37</td>
</tr>
<tr>
<td>RDF</td>
<td>0.03</td>
<td>0.26</td>
<td>1.03</td>
<td>0.62</td>
<td>- 1.72</td>
</tr>
<tr>
<td>NDS</td>
<td>-0.07</td>
<td>0.11</td>
<td>0.93</td>
<td>0.75</td>
<td>- 1.16</td>
</tr>
<tr>
<td>MPRDF</td>
<td>0.08</td>
<td>0.07</td>
<td>1.08</td>
<td>0.95</td>
<td>- 1.23</td>
</tr>
<tr>
<td>MMPI-2-RF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fp-r</td>
<td>0.19</td>
<td>0.18</td>
<td>1.21</td>
<td>0.85</td>
<td>- 1.72</td>
</tr>
<tr>
<td>F-s</td>
<td>0.38*</td>
<td>0.19</td>
<td>1.47</td>
<td>1.01</td>
<td>- 2.14</td>
</tr>
<tr>
<td>FBS-r</td>
<td>-0.01</td>
<td>0.12</td>
<td>0.99</td>
<td>0.78</td>
<td>- 1.25</td>
</tr>
<tr>
<td>RBS</td>
<td>0.39**</td>
<td>0.15</td>
<td>1.47</td>
<td>1.10</td>
<td>- 1.96</td>
</tr>
</tbody>
</table>

Note: SI = standard instructions; OR = feigning instructions; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias Scale; NIM = Negative Impression Management; MAL = Malingering Index; RDS = Roger's Discriminant Function; NDS = Negative Distortion Scale; MPRDF = Malingered Pain-Related Disability Function. B = unstandardized beta coefficients; SE = standard error; CI = confidence Interval; *p < .05; **p < .01; ***p < .001.
Table 7. Hierarchical Logistic Regression Analysis comparing the PVT and MMPI-2-RF scales in differentiating SI from OR groups.

<table>
<thead>
<tr>
<th>Validity Index</th>
<th>Block 1</th>
<th></th>
<th></th>
<th>Block 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
<td>Odds Ratio 95% C.I.</td>
<td>B</td>
<td>SE B</td>
<td>Odds Ratio 95% C.I.</td>
</tr>
<tr>
<td>PVT REY</td>
<td>-0.24</td>
<td>0.18</td>
<td>0.79 0.79 - 0.56 0.56 - 1.12</td>
<td>-0.17</td>
<td>0.32</td>
<td>0.84 0.84 - 0.45 0.45 - 1.57</td>
</tr>
<tr>
<td>RDS</td>
<td>-0.43**</td>
<td>0.13</td>
<td>0.65 0.65 - 0.51 0.51 - 0.83</td>
<td>-0.50*</td>
<td>0.20</td>
<td>0.61 0.61 - 0.41 0.41 - 0.90</td>
</tr>
<tr>
<td>VSVT</td>
<td>-0.16***</td>
<td>0.04</td>
<td>0.85 0.85 - 0.79 0.79 - 0.91</td>
<td>-0.13**</td>
<td>0.04</td>
<td>0.88 0.88 - 0.81 0.81 - 0.95</td>
</tr>
<tr>
<td>MMPI-2-RF F-r</td>
<td></td>
<td></td>
<td></td>
<td>-0.16</td>
<td>0.13</td>
<td>0.86 0.86 - 0.66 0.66 - 1.11</td>
</tr>
<tr>
<td>Fp-r</td>
<td>0.31</td>
<td>0.22</td>
<td>1.36 1.36 - 0.88 0.88 - 2.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fs</td>
<td>0.18</td>
<td>0.23</td>
<td>1.20 1.20 - 0.77 0.77 - 1.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBS</td>
<td>0.00</td>
<td>0.13</td>
<td>1.00 1.00 - 0.78 0.78 - 1.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBS</td>
<td>0.38*</td>
<td>0.17</td>
<td>1.46 1.46 - 1.04 1.04 - 2.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reverse Entry

| MMPI-2-RF F-r  |         |         |                   | -0.16   | 0.13    | 0.86 0.86 - 0.66 0.66 - 1.11 |
| Fp-r           | 0.13    | 0.15    | 1.14 1.14 - 0.85 0.85 - 1.53 |
| Fs             | 0.37*   | 0.17    | 1.45 1.45 - 1.05 1.05 - 2.00 |
| FBS            | -0.04   | 0.10    | 0.96 0.96 - 0.79 0.79 - 1.17 |
| RBS            | 0.41**  | 0.14    | 1.50 1.50 - 1.15 1.15 - 1.97 |
| PVT REY        | -0.17   | 0.32    | 0.84 0.84 - 0.45 0.45 - 1.57 |
| RDS            | -0.50*  | 0.20    | 0.61 0.61 - 0.41 0.41 - 0.90 |
| VSVT           | -0.13** | 0.04    | 0.88 0.88 - 0.81 0.81 - 0.95 |

Note: SI = standard instructions; OR = feigning instructions; REY FIT = Rey Fifteen Item Test; RDS = Reliable Digit Span; VSVT = Victoria Symptom Validity Test; F-r = Infrequent Responses; Fp-r = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias Scale; B = unstandardized beta coefficients; SE = standard error; CI = confidence Interval; *p < .05; **p < .01; ***p < .001.
Table 8.
Hierarchical Logistic Regression Analysis comparing the PVTs and PAI scales and indices in differentiating SI from OR groups.

<table>
<thead>
<tr>
<th>Validity Index</th>
<th>Block 1</th>
<th></th>
<th>Block 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
<td>Odds Ratio</td>
<td>95% C.I.</td>
</tr>
<tr>
<td>PVT REY</td>
<td>-0.24</td>
<td>0.18</td>
<td>0.79</td>
<td>0.56 - 1.12</td>
</tr>
<tr>
<td>RDS</td>
<td>-0.43**</td>
<td>0.13</td>
<td>0.65</td>
<td>0.51 - 0.83</td>
</tr>
<tr>
<td>VSVT</td>
<td>-0.16***</td>
<td>0.04</td>
<td>0.85</td>
<td>0.79 - 0.91</td>
</tr>
<tr>
<td>PAI NIM</td>
<td></td>
<td></td>
<td>0.45***</td>
<td>0.12</td>
</tr>
<tr>
<td>MALI</td>
<td>-0.60</td>
<td>0.42</td>
<td>0.55</td>
<td>0.24 - 1.26</td>
</tr>
<tr>
<td>RDF</td>
<td>0.15</td>
<td>0.35</td>
<td>1.16</td>
<td>0.59 - 2.28</td>
</tr>
<tr>
<td>NDS</td>
<td>0.03</td>
<td>0.15</td>
<td>1.03</td>
<td>0.76 - 1.38</td>
</tr>
<tr>
<td>MPRDF</td>
<td>0.09</td>
<td>0.10</td>
<td>1.09</td>
<td>0.91 - 1.32</td>
</tr>
</tbody>
</table>

Reverse Entry

| PAI NIM        | 0.49*** | 0.10 | 1.63 | 1.36 - 1.97 | 0.45*** | 0.12 | 1.57 | 1.23 - 2.00 |
| MALI           | -0.29   | 0.31 | 0.75 | 0.41 - 1.37 | -0.60   | 0.42 | 0.55 | 0.24 - 1.26 |
| RDF            | 0.03    | 0.26 | 1.03 | 0.62 - 1.72 | 0.15    | 0.35 | 1.16 | 0.59 - 2.28 |
| NDS            | -0.07   | 0.11 | 0.93 | 0.75 - 1.16 | 0.03    | 0.15 | 1.03 | 0.76 - 1.38 |
| MPRDF          | 0.08    | 0.67 | 1.08 | 0.95 - 1.23 | 0.09    | 0.10 | 1.09 | 0.91 - 1.32 |

Note: SI = standard instructions; OR = over-reporting instructions; REY FIT = Rey Fifteen Item Test; RDS = Reliable Digit Span; VSVT = Victoria Symptom Validity Test; F-r = Infrequent Responses; Fp = Infrequent Psychopathology Responses; Fs = Infrequent Somatic Responses; FBS-r = Symptom Validity; RBS = Response Bias Scale. B = unstandardized beta coefficients; SE = standard error; CI = confidence Interval; *p < .05; **p < .01; ***p < .001.
Appendix A

Post Experimental Questionnaire.

1. What were you asked to do in this study?
   a) Answer all paper pencil questions and perform on all tests honestly and to the best of my ability.
   b) Answer all paper pencil questions and perform on all tests as though I have a mild traumatic brain injury.
   c) Answer all questions as though I have a mild traumatic brain injury, but perform to the best of my ability on the tests.
   d) Answer all paper pencil questions honestly and perform on all tests as though I have a mild traumatic brain injury.

2. How certain are you that you answered questions and performed on all tests based on the instructions you were provided?

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Uncertain</td>
<td>Neither Certain or Uncertain</td>
<td>Certain</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

MTBI and related symptoms description provided to over-reporting instruction participants.

A mild traumatic brain injury (mTBI) typically follows an event where you hit your head. It can result in many kinds of physical, cognitive and behavioral / emotional impairments. These symptoms, when they persist, can interfere with an individuals’ ability to engage in various daily living activities, such as work and completing school assignments. While the specific constellation of symptoms that can result following a mTBI can vary, there are some commonalities. These can include:

Physical
- movement problems
- coordination and balance problems
- dizziness
- chronic pain
- fatigue
- seizures
- sleep problems

Sensory
- impairment of sense of smell and/or taste
- vision or hearing difficulties

Communication
- impaired speech output
- word-finding problems
- difficulty understanding oral, written or non-verbal language

Emotional
- difficulty expressing emotions appropriately
- mood swings
- irritability
- anxiety
- depression

Cognitive
- slowed information processing
- memory loss
- memory-processing (e.g., learning) problems
- problems with concentration and attention
- impaired judgment
- problem-solving difficulties
- social behavioural problems
Appendix C

Case vignette and simulation instructions provided to over-reporting participants

“The purpose of this experiment is to see if individuals can successfully fake symptoms associated with a condition known as mild traumatic brain injury (mTBI). Before I give you your instructions about what you are to do in this experiment, I want you to study and familiarize yourself with the symptoms of mTBI. I’ll give you five minutes to review the following sheet of paper, which outlines briefly, but in detail, the symptoms associated with mTBI.

Now I'm going to give you some instructions about what your tasks are in this experiment. This includes a scenario where you might find yourself in a position to fake symptoms of mTBI. Imagine, for example, you are doing poorly in a course and you need more time to complete course work. You need an extension. You have just learned about mTBI in one of your psychology courses and you have decided to fake this condition in order to get an extension to complete your coursework. You visit your family doctor in order to get a medical certificate documenting that you are suffering from mTBI. You report to your doctor that you slipped and fell and hit your head. You tell your doctor that subsequent to that fall you are having difficulties with thinking clearly. You tell your doctor that this has affected your ability to complete your coursework. Before your doctor will issue the medical certificate, he/she refers you to a specialist – a clinical neuropsychologist – who will administer a number of tests to determine if you have symptoms consistent with mTBI. Keeping in mind the definition and symptoms of mTBI that you have just read and studied, your task in this experiment is to complete the following tests as if you are suffering from mTBI.

It is important to note that these tests include questions and cognitive tasks related to mTBI symptoms. It is equally important to note that these tests are designed in such a way to detect if you are exaggerating your symptoms to the extent that they are unbelievable. Do you understand what your task is?”