

Summer 2018

Latent Class Analysis of Malingering Classifications Using Performance and Symptom Validity Measures in a Civil Forensic Setting

Willie Floyd McBride III
Old Dominion University, williebeamen89@gmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_etds



Part of the [Clinical Psychology Commons](#)

Recommended Citation

McBride, Willie F.. "Latent Class Analysis of Malingering Classifications Using Performance and Symptom Validity Measures in a Civil Forensic Setting" (2018). Doctor of Philosophy (PhD), Dissertation, Psychology, Old Dominion University, DOI: 10.25777/g5kr-z765
https://digitalcommons.odu.edu/psychology_etds/80

This Dissertation is brought to you for free and open access by the Psychology at ODU Digital Commons. It has been accepted for inclusion in Psychology Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

LATENT CLASS ANALYSIS OF MALINGERING CLASSIFICATIONS USING
PERFORMANCE AND SYMPTOM VALIDITY MEASURES IN A CIVIL FORENSIC

SETTING

by

Willie Floyd McBride III
Bachelor of Arts, May 2011, University of Louisville
Master of Science, May 2013, Eastern Kentucky University

A Dissertation Submitted to the Graduate Faculties of
Eastern Virginia Medical School
Norfolk State University
Old Dominion University
In Partial Requirement for the Degree of

DOCTOR OF PHILOSOPHY

CLINICAL PSYCHOLOGY

VIRGINIA CONSORTIUM PROGRAM IN CLINICAL PSYCHOLOGY
August 2018

Approved by:

Richard W. Handel (Director)

Jennifer M. Flaherty (Member)

Michael L. Stutts (Member)

James M. Henson (Member)

Roger Gervais (Member)

ABSTRACT

LATENT CLASS ANALYSIS OF MALINGERING CLASSIFICATIONS USING PERFORMANCE AND SYMPTOM VALIDITY MEASURES IN A CIVIL FORENSIC SETTING

Willie Floyd McBride III
Eastern Virginia Medical School, 2018
Director: Dr. Richard W. Handel

The study examined the applicability of various validity measures and their ability to identify patterns of invalid responding (i.e., malingering) in an archival sample of civil forensic litigants. A latent clustering approach was used to create profiles comprised of the response patterns from validity measures of different domains including three performance validity tests (PVT), five symptom validity scales (SVT), and an embedded validity indicator (EVI). Latent class analysis (LCA) was used to enumerate subgroups of malingering and to assess differences between the subgroups.

Results demonstrated five distinct classes. The classes revealed complex patterns of symptom endorsement and performance on validity measures that were indicative of malingering. The profiles were labeled as follows: *Definite Malingering*, *Probable Neurocognitive Malingering*, *Probable Symptom Malingering*, *Possible Malingering*, and *Valid Responders*. The *Definite Malingering* class had the highest failure rates for performance (below cut-off and below chance) and embedded validity test, as well as, the largest percentage of symptom over-endorsement and invalid responding on the symptom validity scales [i.e., F-r, Fp-r, Fs, FBS-r, RBS of the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF)]. The *Probable Neurocognitive Malingering* class had the second highest percentage of below cut-off failure (below chance performance was absent from this group) on the performance validity measures but had less evidence of invalid responding on the symptom

validity scales as opposed to the *Definite Malingering* or *Probable Symptom Malingering* classes. The *Probable Symptom Malingering* class had a higher percentage of PVT above cutoff performance (i.e., pass), however, it had a large percentage of symptom over-endorsement and invalid responding second only to the *Definite Malingering* class. The *Possible Malingering* class demonstrated minimal evidence of failure on the three PVTs (i.e., less than 15% failed below cut-off) and there was overall less evidence of symptom over-reporting and invalid responding. The *Valid Responders* class had the highest level of passing on the performance validity tests and had the highest overall percentage of valid responding on the symptom validity scales. Implications of the findings and limitations are discussed.

Copyright, 2018, by Willie Floyd McBride III, All Rights Reserved.

This dissertation is dedicated to Ida Mae McBride, my Nanny. I could never repay you for the unconditional love you showed a little boy that stole your heart in that hospital room.

ACKNOWLEDGEMENTS

I would like to give thanks to those individuals who have had a hand in helping me succeed throughout my graduate school career and the people who played a part in the culmination of my doctoral degree. First, I would like to thank my advisor, Dr. Richard W. Handel, for his support and guidance throughout this arduous doctoral process.

Next, I would also like to thank those individuals that served on my dissertation committee, including Drs. Jennifer Flaherty, Michael Stutts, Matt Henson, and Roger Gervais. I would especially like to thank Dr. Roger Gervais who provided data for this project and extending his support and expertise to an unknown doctoral student.

I would like to thank my friends, colleagues, and roommates, future Drs. Kristoffer Park, Stefan LaTulip, and Charles Freligh who have helped to keep me grounded and focused while surviving graduate school. To the future Dr. Randy Whestone, I thank you for being my best friend and a brother to me. Your love, our friendship, fellowship, and brotherhood have helped me through this process immensely.

I would also like to thank my mentors, Drs. Dustin B. Wygant and Robert P. Granacher. Seven years ago, I walked into both of their offices, and since then they have served as guides and they have invested in my growth to help move forward my professional career and personal life. For this I am eternally grateful.

I would like to thank my family, who have supported me through every phase of my life. To my mother and father, Latricia and Willie, who far from perfect, have shown me unconditional love that has allowed me to pursue my dreams. To my grandparents, Ida and Willie McBride Sr., you both are wonders and have instilled in me an “old soul” that will never

fade. To Inez and James “Teddy” Powell, although you are not here with me on this earth, I know you are forever my angels watching over. To my sister, Rockia “my booby,” I could not have asked for a better sister and I thank you for your love and “MJ” shenanigans with me.

Finally, to Simone, I believe that you were brought into my life for a very special reason. You are here to challenge, inspire, love, and support me. I could never repay you for dealing with my frustrating moments, my lowest moments, and my highs.

TABLE OF CONTENTS

	Page
LIST OF TABLES	x
LIST OF FIGURES	xi
 Chapter	
I. INTRODUCTION	1
RESPONSE STYLES.....	4
MALINGERING AS DIAGNOSTIC CRITERIA.....	6
PREVALENCE OF MALINGERING.....	6
ALTERNATIVE MALINGERING CRITERIA	7
DOMAINS OF MALINGERING	13
MALINGERING OF PSYCHOPATHOLOGY	13
NEUROCOGNITIVE MALINGERING.....	15
SOMATIC MALINGERING.....	17
ACCURACY OF DETECTION STRATEGIES	17
RESEARCH DESIGNS IN MALINGERING.....	19
ASSESSMENT MEASURES OF MALINGERING.....	20
PERFORMANCE VALIDITY TESTING (PVT)	21
SYMPTOM VALIDITY TESTING (SVT).....	24
EMBEDDED PERFORMANCE VALIDITY INDICATORS WITHIN STANDARDIZED NEUROPSYCHOLOGICAL MEASURES	27
MMPI-2-RF AND MALINGERING	28
MALINGERING AS A CATEGORICAL OR CONTINUOUS CONSTRUCT	33
LATENT CLASS ANALYSIS	36
THE PRESENT STUDY	38
II. METHOD	39
PARTICIPANTS.....	39
INSTRUMENTS AND MEASURES	44
PERFORMANCE VALIDITY TESTS (PVT).....	44
SYMPTOM VALIDITY TEST (SVT).....	47
EMBEDDED PERFORMANCE VALIDITY INDICATOR (EVI).....	47
CLASSIFICATION OF VALIDITY MEASURES FOR LATENT CLASS ANALYSIS.....	48
ANALYSES.....	52
III. RESULTS.....	55
DESCRIPTIVE STATISTICS	55
THE LATENT CLASSES	61
IV. DISCUSSION.....	67

	Page
IMPLICATIONS.....	70
LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH.....	73
V. CONCLUSIONS.....	76
REFERENCES.....	77
VITA.....	102

LIST OF TABLES

Table	Page
1. Criteria for Definite, Probable, and Possible Malingered Neurocognitive Dysfunction ...	12
2. Assessment Type	41
3. Accident Type	42
4. Referral Type.....	42
5. Diagnoses	43
6. Scoring Criteria for Performance Validity Test (PVT) Performance.....	49
7. Scoring Criteria for Embedded Performance Validity Indicator (EVI).....	50
8. Scoring Criteria for Symptom Validity Test (SVT) Endorsement on the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF)	51
9. Total and Percentages of PVT Performance by Levels	55
10. Total and Percentages of EVI Performance by Levels	55
11. Total and Percentages of SVT Endorsement by Levels.....	56
12. Model Fit Information for 1-7 Class LCA Models.....	58
13. Final Class Counts and Proportions for the Latent Classes Based on their Most Likely Latent Class Membership.....	59
14. Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column)	60
15. Classification Probabilities for Most Likely Latent Class Membership by Latent Class.....	60
16. Conditional Item Probabilities for Five Class Latent Model.	64
17. Number of Criteria Failed and/or Over-endorsed for the Five Class Model	66

LIST OF FIGURES

Figure	Page
1. Conditional Item Probability Profile Plot for Five Class Model.....	65

CHAPTER I

INTRODUCTION

In the context of forensic psychological evaluations, the veracity of a client's presentation is one of the most important questions. This question is particularly important when an incentive (internal or external) is present. External influences on a patient's presentation can include financial or material incentives, or shirking work and/or school responsibilities. Internal influences may include reactions to one's credibility being questioned, negative perceptions related to having a mental disorder or disability status, the consequences of psychopathology, or the desire to obtain underserved benefits (Rogers & Bender, 2003). One hypothetical scenario that could occur in forensic assessments, involve individuals presenting for an evaluation after a minor motor vehicle accident caused by another party. It may be in that individual's interest to present as impaired as possible, both physically and mentally, in order to maximize the potential financial gains despite receiving only minor injuries. Therefore, these individuals have the potential to exaggerate their symptoms or report symptoms that are not actually present (Kane, 1999). Deception, alternatively defined as dissimulation, occurs when an individual misrepresents and/or fabricates (i.e., engages in response bias) his or her clinical presentation for potential gain. The potential for deception has overarching effects as the range of legal decisions that are impacted are broad and vary across civil, criminal, and family issues (Heilbrun, 1992). Additionally, this type of deception places a large financial burden on the health care system (Bush & Graver, 2013).

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; APA, 2013) defines these characteristics of misrepresentation as malingering. Malingering is the intentional production of false or grossly exaggerated physical or psychological symptoms

motivated by external incentives. To further complicate the understanding of misrepresentation and malingering, individuals are capable of feigning a diversity of symptoms including psychological, cognitive, and somatic complaints.

Researchers have worked to categorize various types of dissimulation and malingering to further understand its influence in forensic assessment (Ben-Porath, 2003; Rogers, 2018). Numerous psychological assessment measures have been developed that directly and/or indirectly address response bias. Broad self-report inventories such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989; Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) and the more recent, MMPI-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) contain validity scales designed to detect various response styles. Structured clinical interviews, such as the Structured Interview of Reported Symptoms (SIRS; Rogers Bagby, & Dickens., 1992), and its second edition (SIRS-2; Rogers, Sewell, Gillard, 2010) assess feigning and related response styles, also. Stand-alone task engagement measures have been developed to detect the feigning of cognitive impairment, such as the Test of Memory Malingering (TOMM; Tombaugh, 1996), Word Memory Test (WMT; Green, 2003), Medical Symptom Validity Test (MSVT; Green, 2004), or the Nonverbal Medical Symptom Validity Test (NV-MSVT; Green, 2008). Furthermore, many domain-specific cognitive and intelligence tests that are given during the course of a standard neuropsychological evaluation contain embedded performance validity indicators that detect poor response bias (i.e., poor task engagement).

These measures have allowed researchers to closely examine the effectiveness of response style detection strategies at capturing malingering, both in clinical and forensic settings. To examine their effectiveness simulation designs, known-groups comparisons, and differential

prevalence designs have predominated the empirical study of malingering (Rogers, 2018). While these research designs have their strengths, weaknesses still remain including generalizability, the establishment of true known groups based on accurate assessment, and the inability to establish prevalence rates among groups.

A determination of malingering carries significant consequences. Labeling someone as a malingering who is, in fact, truthful can have dire implications for that individual and those around them. Conversely, failing to identify individuals who are misrepresenting their symptoms can create an even greater burden on the legal and health care system. Given the ramifications, it is vital for clinicians who perform these evaluations to have a level of certainty that minimizes the likelihood for misdiagnosis. With this in mind, debates have ensued about whether malingering is a dichotomous (i.e., malingering or honestly responding) or a dimensional construct (e.g., possible, probable, or definite malingering). For these reasons, numerous researchers have advocated for a multi-method approach to assessing for malingering by requiring the use of multiple validity measures, from different domains, to make a final determination (Boone, 2008).

Latent class analysis is a more recent technique that has been applied in social and behavioral sciences (Larrabee, 2012). Latent class analysis (LCA) works to identify unobserved (latent) categorical or continuous variables (latent constructs) that account for the covariance between two or more observed (manifest) variables (Thomas, Lanyon, and Millsap, 2009). Therefore, this technique has potential utility for understanding malingering.

This study sets to examine the latent structure of malingering and to determine if various validity measures of multiple domains can accurately capture groups of malingering. LCA can be used to evaluate the classification accuracy of multiple measures of malingering (symptom,

performance validity, and embedded validity indicators) and determine the presence of various subgroups of malingering. Chapter One presents a review of the literature on malingering and response bias, the different domains of malingering, and the methods for assessing malingering. Next, various research methodologies for evaluating malingering-detection instruments will be examined, with an emphasis on symptom, performance validity, and embedded validity indicators. In Chapter Two, the sample and instruments used in the study are detailed. A rationale for the proposed study and set of goals are presented. The analysis for evaluation will additionally be discussed. Chapter Three presents the results of the study. Chapter Four discusses the findings their implications, along with, limitations and future research directions.

Response Styles

Rogers (2018) details six major response styles that focus primarily on malingering and its associated response patterns. *Malingering* refers to the intentional fabrication or gross exaggeration of physical and/or psychological symptoms for an external incentive.

Defensiveness, as opposed to malingering, represents a guarded approach or minimization of symptoms and/or problems. *Irrelevant* and its subset, *random responding*, occur when an individual does not attend to the assessment process. *Honest responding* describes an accurate portrayal of an individual's symptoms, whereas *hybrid responding* refers to a mixture of response styles.

The presence of numerous response styles exemplifies the difficulty in classifying and assessing for malingering. The assessment process becomes increasingly difficult when individuals misrepresent symptoms across psychological, physical, and/or cognitive domains. Due to the potential for a comorbid symptom presentation, self-report measures represent a useful means for detecting symptoms across multiple domains. Ben-Porath (2003) details

concerns regarding malingering on self-report measures that impact validity: *over-reporting* and *under-reporting*. Described as protocol validity, this concept refers to the truthfulness of the results an individual produces on a test administration. Threats to protocol validity include *non-content* and *content based invalid responding* (Ben-Porath, 2003). Of greater importance to this study, *content based invalid responding* occurs when a test taker distorts his or her responses to create differing presentations including the over and under-reporting of symptoms. Ben-Porath (2012) describes *over-reporting* as synonymous with feigning, faking bad, and negative response bias. *Over-reporting* occurs when individual answer in a way to portray problems he or she does not truly have or exaggerate the significance of difficulties they are experiencing at that point in time. Intentional over-reporting involves an individual knowingly shifting their responses to appear more impaired. *Under-reporting* refers to a pattern of positive impression management associated with presenting in an overly positive light or without major faults (e.g., faking good or positive malingering; Ben-Porath, 2012). Over-reporting symptoms, however, does not necessarily lead to a an automatic conclusion of malingering, as there may be various reasons for one to over-report symptoms (e.g., psychological distress, careless responding, and/or feigning). Malingering occurs when motivations are INTENTIONAL and influenced by EXTERNAL gains (APA, 2013).

Historically, concerns regarding protocol validity have been highlighted by Allport (1937) and Ellis (1946). They theorized that dependence on individual self-report to generate an accurate self-portrayal would be its greatest challenge. To resolve this dilemma, many self-report measures contain protocol validity scales to assess various content-based response styles. Examples of these measures include the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943), its revisions, the MMPI-2 (Butcher et al., 1989; Butcher et al.,

2001) and MMPI-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008). Despite the inclusion of validity scales in many self-report measures to assess over-reporting, these measures are unable to determine the motivation for the response bias. Therefore, additional information is necessitated to reach a determination of malingering.

Malingering as Diagnostic Criteria

The American Psychiatric Association (2013) groups malingering in a section of the DSM-5 that discusses various conditions and problems that are a focus of clinical attention or that may otherwise affect the diagnosis, course, prognosis, or treatment of a patient's mental disorder. Therefore, these conditions are not considered mental disorders (i.e., malingering is not a diagnosis but a determination) and receive a corresponding V or Z code according to the International Classification of Disease, 10th Revision, Clinical Modification (ICD-10-CM; PLCH & McCormack, 2010). Accordingly, malingering is defined as the intentional fabrication of false or grossly exaggerated physical symptoms. Furthermore, the motivation for such exaggerations must be caused by external incentives which differentiate malingering from somatic symptom disorders, such as factitious or conversion disorder (internal incentives). The DSM-5 also suggests that malingering should be considered under the following conditions: medicolegal context; marked discrepancies between the patients' claims and objective findings; the patient is uncooperative during a diagnostic evaluation or in compliance with treatment; or antisocial personality traits.

Prevalence of Malingering

Despite considerable research being devoted to examining malingering, there remains uncertainty regarding the exact prevalence or base rates of malingering. In medicolegal settings, the detection of malingering is crucial. Larrabee (2003) reviewed findings from 11 studies that

reported information related to malingering base rates. Of 1,363 individuals in medicolegal evaluations, 40% presented with profiles suggestive of malingering. Across diagnostic domains, differences in base rates of malingering have also been observed. 25-30% of disability claimants with diagnoses such as fibromyalgia, chronic fatigue, or major depressive disorder demonstrated signs of malingering on forced-choice tests (Gervais, Russel, Green, Allen, Ferrari, & Pieschl, 2001; Green, Rohling, Lees-Haley, & Allen, 2001; Van der Werf, Prins, Jongen, van der Meer, & Bleijenberg, 2000). Gervais and colleagues (2001) reported approximately 40% of chronic pain disability claimants performed in a manner suggestive of malingering. Findings from personal injury litigation, worker's compensation, or disability claim evaluations reported 25% to 30% of cases demonstrating probable malingering when measures such as the MMPI-2 or forced choice cognitive tests were used (Green, Rohling, Lees-Haley, & Allen, 2001; Lees-Haley, 1997). Base rates of malingering in criminal forensic settings demonstrated similar findings of 20% to 30% (Frederick, 2000; Miller & Rohling, 2001; Rogers, 2018). Regarding head injury compensation seeking individuals, 37% produced profiles suggesting malingering when the combination of clinical criteria and forced choice cognitive measures were considered (Greiffenstein, Baker, & Gola, 1994). Mittenberg, Patton, Canyock, and Condit (2002) surveyed 144 clinicians in neuropsychological practice to approximate the percentage of cases they evaluated that suggested probable symptom exaggeration or malingering. 8% to 31% of the cases suggested probable malingering or symptom exaggeration.

Alternative Malingering Criteria

Numerous researchers have advocated for greater uniformity within the diagnostic criteria for malingering. Although the DSM-5 has established malingering criteria useful for clinical practice, weaknesses remain (Larrabee 2012; Rogers, 2018). The V-Code status indicates

that the diagnosis lacks formal diagnostic criteria which hinder its usefulness in everyday neuropsychological assessment. Furthermore, the DSM criteria include the component of “volitional exaggeration,” which proves difficult to determine in clinical practice. Additionally, the distinctions between malingering, factitious disorder, and conversion disorder complicate differential diagnostic capabilities (Rogers, 2018). Factitious disorder, like malingering, involves intentional production of symptoms suggestive of injury or disease. However, factitious disorder is characterized by internal motivations that are psychological in nature and have less to do with an external incentive. Another disorder that complicates the determination of malingering is conversion disorder. Conversion disorder is defined by altered motor or sensory functions that are incompatible with recognized neurological or medical conditions (Larrabee, 2012). Symptoms of conversion disorder are said not to be under volitional control but are psychological in nature. Another impediment regarding the DSM-5 criteria concerns the inability to rule in malingering if factitious or conversion disorders are present. As stated by Slick, Sherman, and Iverson (1999), comorbidity of malingering and factitious disorder cannot occur in DSM-IV. Lastly, the DSM-5 provides minimal assistance on the assessment of malingering in the context of neuropsychological evaluations. Several notable researchers (Greiffenstein et al., 1994; Rogers, 1990a, 1990b), though, have worked to develop objective criteria for malingering.

Rogers (2018) proposed a classification model of malingering that incorporated four detection strategies: *over-endorsement of symptoms rarely endorsed by patients with mental illness*; *an endorsement of a large array of symptoms*; *over-endorsement of blatant symptoms*; and the *endorsement of unusual or fantastical symptoms*. His malingering criteria were developed within the context of a psychiatric assessment and therefore, did not address specific issues related to neuropsychological malingering.

Greiffenstein and his colleagues (1994) established a set of criteria for malingering of memory dysfunction in the context of a neuropsychological evaluation. Malingering was considered when 2 or more of the following criteria were met: (1) two or more severe impairment ratings on neuropsychological tests; (2) an unrealistic history of symptoms; (3) complete disability in work or social roles; or (4) reports of remote memory loss. These criteria lacked the external incentive criterion and were focused primarily on demonstrating objective grouping criteria for malingering that were uncontaminated by incentive criteria. Furthermore, they suggest two significant problems with the litigation status (i.e., external incentive): failure to find non-litigation minor head injury patients as controls, and failure to consider other incentives for manufacture disability (e.g. wage replacement benefits, avoidance of responsibility, and sympathy). Slick et al. (1999) stated that the Greiffenstein criteria lacked concreteness regarding an explicit operational definition of malingering as well as a lack of reference for malingering in other neurocognitive domains.

Specific malingering criteria related to neurocognitive dysfunction have since been developed to detect feigning and improve on previously published malingering criteria (Slick, Sherman, & Iverson, 1999). Malingered Neurocognitive Dysfunction (MND; Slick et al., 1999) refers to the intentional exaggeration of cognitive dysfunction for an external incentive. Additionally, the MND criteria have been used in numerous studies (e.g., Ardolf, Denney, & Houston, 2007; Bianchini, Greve, & Love, 2003; Bianchini, Love, Greve, & Adams, 2005; Curtis, Greve, Bianchini, & Brennan, 2006; Etherton, Bianchini, Greve, & Heinly, 2005; Greve et al., 2006; Greve, Bianchini, & Doane, 2006; Greve et al., 2009; Greve, Bianchini, Love, Brennan, & Heinly, 2006; Greve, Bianchini, Mathias, Houston, & Crouch, 2002, 2003; Greve et

al., 2007; Heinly, Greve, Bianchini, Love, & Brennan, 2005; Larrabee, 2003b; Mathias, Greve, Bianchini, Houston, & Crouch, 2002).

Summarily, the MND criteria require consideration of various conditions such as (A) the presence of an external incentive, (B) consideration of evidence from neuropsychological testing, (C) self-report assessment results, and (D) that conditions meeting the criteria for B and C are not attributed to psychiatric, developmental, and/or neurological factors. Evidence for Criteria B (neuropsychological testing evidence) can be captured through 6 factors. One condition includes (B1) *definite negative response bias*, which is defined as below chance performance ($p < .05$) on one or more forced-choice measures of cognitive function. (B2) *Probable response bias* indicates a performance consistent with feigning on one or more well-validated psychometric tests or indices. Evidence concerning discrepant patterns includes marked deviations between test data and known patterns of brain functioning (B3), test data and observed behaviors (B4), test data and reliable collateral reports (B5), and last, discrepancies between test data and documented background history (B6).

Criteria C conditions for evidence on self-report measures include a self-report history discrepant with documented history (C1), self-report symptoms are discrepant with regards to known patterns of brain functioning (C2), behavioral observations (C3), or information obtained from collateral informants (C4). Last, evidence of exaggerated or fabricated psychological dysfunction is enough to constitute evidence for criteria C (C5).

Additionally, Slick and his colleagues (1999) detail specifications regarding various gradations of malingering: *possible*, *probable*, and *definite malingering*. *Definite Malingered Neurocognitive Dysfunction (MND)* is defined by the presence of clear and compelling evidence of willful exaggeration or fabrication of cognitive dysfunction and the plausible alternative

explanations which include the presence of a substantial external incentive (criterion A), *definite negative response bias* (criterion B1), and that criterion B is not caused by psychiatric, neurological, or development factors (criterion D).

Probable MND is defined by the presence of evidence strongly suggesting volitional exaggeration or fabrication of cognitive dysfunction in the absence of plausible alternative explanations. The criteria include a substantial external incentive (criterion A), two or more types of evidence from neuropsychological testing, excluding definite response bias (i.e., two or more of Criteria B2-B6). Alternatively, this criterion can be demonstrated through one type of neuropsychological testing, excluding definite negative response bias, and one or more types of evidence from Self-Report (i.e., one of Criteria B2-B6 and one or more of Criteria C1-C5). Criteria D constitute the final component for *Probable MND*.

Possible MND is characterized by evidence suggesting intentional exaggeration of cognitive dysfunction and the absence of plausible alternative explanations. Additionally, the criteria of Definite or Probable MND may be met with the exception that criteria D cannot be ruled out. Therefore, there is evidence for substantial external incentive (Criterion A), evidence from self-report (one or more of C1-C5), and criteria D are met.

Table 1 contains a summary of the criteria for Malingered Neurocognitive Dysfunction adapted from the Slick et al. (1999) criteria.

Table 1

*Criteria for Definite, Probable, and Possible Malingered Neurocognitive Dysfunction.***Definite MND**

1. Presence of a substantial external incentive [Criterion A]
2. Definite negative response bias [Criterion B1]
3. Behaviors meeting necessary criteria from criterion B are not fully explained by psychiatric, neurological, or developmental factors [Criterion D]

Probable MND

1. Presence of a substantial external incentive [Criterion A]
2. Two or more types of evidence from neuropsychological testing except for Criterion B1 [two or more of Criteria B2-B6]

Or

One piece of evidence from neuropsychological testing, excluding Criterion B1, and one or more types of evidence from Self-Report [one of Criterion B2-B6 and one or more of Criteria C1-C5]

3. Behaviors meeting necessary criteria from criterion B are not fully explained by psychiatric, neurological, or developmental factors [Criterion D]

Possible MND

1. Presence of a substantial external incentive [Criterion A]
2. Evidence from type of evidence from neuropsychological testing except for Criterion B1 [one of Criteria B2-B6]

Or

Evidence from Self-report [one of Criteria C1-C5]

Note. Adapted from Slick et al. (1999) Malingered Neurocognitive Dysfunction criteria.

Domains of Malingering

According to Rogers (2018), malingering occurs through three broad domains: psychological, neurocognitive, and physical/medical symptoms. Much of the literature has focused on developing detection strategies for psychological and neurocognitive symptoms with fewer studies devoted to physical/medical symptoms.

Malingering of Psychopathology

Specific to psychological malingering, individuals who feign in this context report a broad array of psychological dysfunction, often, in an unsophisticated manner. Detection strategies involve eight indicators, according to Rogers and Bender (2003). The detection strategies center on the individual having a lack of knowledge regarding psychological disorders and demonstrating an inconsistency between their reported symptoms and observed behavior.

Additionally, these individuals endorse very infrequent symptoms that are highly uncommon even in patients with true psychopathology (*Rare symptoms*). Endorsement of these symptoms suggests the individual is unsophisticated (unknowledgeable) regarding the disorder they are attempting to portray. Another psychological malingering strategy regarding infrequent symptom endorsement involves outrageous symptom reports (*Improbable Symptoms*). Increases in this type of endorsement diminishes the credibility of the individuals' self-report.

Additionally, there is a propensity for malingerers to report unusual symptom combinations (*Symptom Combination*) and overestimate most endorsed symptoms as extreme with minimal variation. Furthermore, endorsement of large numbers of symptoms (*Indiscriminant Symptom Endorsement*) is prioritized over selective symptom endorsement strategies. This is also seen in the over-endorsement of obvious symptoms with minimal specificity or elaboration.

Psychological malingering detection strategies rely on observable behaviors as well. The *Erroneous Stereotypes* approach suggests individuals may portray symptoms or behaviors associated with a given disorder but these symptoms and/or behaviors are inaccurate or misconceived. An additional observation strategy, involves the *Reported versus Observed Symptoms* where individuals demonstrate significant discrepancies between their symptom report and clinically observed presentation.

Several self-report measures and structured clinical interviews are equipped to examine each of these malingering detection strategies. Many measures contain validity scales that were developed to capture these response styles. The MMPI-2 and MMPI-2-RF are broad self-report inventories that contain validity scales that assess content non-responsiveness (i.e., over-reporting, and under-reporting). These self-report measures also assess the validity of somatic and cognitive complaints, in addition to psychopathology.

The Structure Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) is a structured interview for the assessment of feigned mental disorders. It was developed specifically to detect feigned psychosis and mental disorders as well as related response styles (Rogers & Bender, 2003). Individuals are classified into three categories based on response styles: feigning, indeterminate, and non-feigning (Rogers, Bagby, & Dickens, 1992). Furthermore, Rogers and Bender (2003) reported that the SIRS is the most validated measure for the assessment of feigned psychopathology. Psychological malingering has received more research attention in comparison to malingering of cognitive and/or somatic symptoms (Lanyon, 2003; Larrabee, 2012).

Neurocognitive Malingering

Symptom self-reports of neurocognitive dysfunction present another area for potential exaggeration in the medicolegal context. In general, injuries resulting in neurocognitive impairments can impact one's executive functioning, memory, attention, language, and processing abilities (Kaufman, Geyer, & Milstein, 2016). Injuries can also result in memory, language, movement, or recognition disorders such as amnesias, aphasias, apraxia, or agnosia. Accurate assessments of these symptoms have serious implications regarding diagnosis, treatment, and in the context of this study, medicolegal restitution.

Advancements in neuroimaging have improved diagnostic capabilities regarding structural causes of neurocognitive dysfunction. Neuroimaging, while advantageous, is limited in its ability to elucidate functional impairments that may be resultant from structural injuries. Therefore, neuropsychological assessment remains a crucial component to objectively verify self-reports of cognitive dysfunction in conjunction with imaging (Boone, 2008; Larrabee, 2012).

Individuals attempting to exaggerate cognitive complaints may report a myriad of symptoms. However, there are several cognitive domains that provide a hierarchy for how these individuals typically present when attempting to demonstrate a brain injury: recognition memory, basic attention, overlearned information, and motor strength/dexterity (Boone, 2013). Therefore, measures of cognitive malingering rely on these domains to identify a significant percentage of malingerers (Rogers, 2018).

Rogers (2018) outlines distinct differences between psychopathology and cognitive feigning. Malingered psychopathology requires the fabrication or exaggeration of symptoms related to a mental disorder (symptom presentation). Conversely, feigners of cognitive impairment only need to demonstrate a lack of knowledge or appear to expend effort, but provide

incorrect responses. For instance, an individual claiming memory impairment performs poorly on measures of memory recall and recognition as a means of verifying their reported impairment. This is applicable to other domains of cognitive functioning as well. This approach to feigned cognitive impairment is characterized as “effortful failure” and requires gross exaggeration of intellectual and neuropsychological deficits for an external goal (Rogers & Bender, 2003).

Rogers and Bender (2003) have outlined how detection strategies are grouped into two domains: results demonstrating *excessive impairment* and *detection by unexpected patterns*. *Excessive impairment* is characterized by failure of very easy items (e.g., floor effect and recognition versus recall) and failures below chance on forced-choice testing. Unexpected patterns are observed when the individual performs differently on easy and difficult items (i.e., performance curve) and unexpected answers on forced-choice formats (i.e., magnitude of error).

Rogers (2018) details six detection strategies for feigned cognitive impairment. Floor effect strategies rationale is based on the idea that malingerers are unable to differentiate which cognitive abilities are unlikely to be compromised in individuals with genuine neurocognitive deficits. *Symptom validity testing* (SVT) examines improbable failure rates based on statistical probability. *Forced-choice testing* (FCT) assesses individuals’ ability to discriminate between two options at a time for multiple items compared to the likelihood of success based on chance alone. The second groups of detection strategies encapsulate unexpected patterns. *Magnitude of error* (MOE) evaluates the degree of inaccuracy for incorrect responses. *Performance curve* is based on the idea that feigners do not take into account item difficulty in choosing which items to fail (Rogers & Bender, 2003). An example of this detection strategy can be seen when an individual performs better on multiplication problems than simple addition and/or subtraction problems. *Atypical performance* suggests that feigners produce deficits in certain cognitive

domains while maintaining fewer deficits in estimates of overall functioning (Rogers, 2018). Severe deficits on simple measures of manual dexterity and tactile sensation are characteristic of this pattern of performance (Binder & Willis, 1991; Heaton, Smith, Lehman, Vogt, 1978). A variation of atypical psychological sequelae involves the report of symptoms of a mental disorder or physical complaints not typically found in individuals with cognitive impairments (Rogers & Binder, 2003). Last, *marked inconsistencies* (Rogers, 2018) involve discrepancies between what is expected after an injury and what is reported or observed on testing, requiring explanation.

Somatic Malingering

Somatic malingering involves symptom exaggeration characterized by the endorsement of physical complaints or pain. Somatic feigning presents a much greater challenge to detect than either malingering in psychological or cognitive domains (Granacher & Berry, 2008). Significantly less research has been devoted to somatic malingering despite the high prevalence of individuals with physical/somatic related disabilities. Disorders such as Somatic Symptom Disorder and Factitious Disorder which have similar clinical presentations are difficult to distinguish from somatic malingering. For these disorders, the motivation for feigning is primarily internal or psychologically driven. Detection measures grounded in physical and somatic symptomology are necessary for distinguishing somatic malingering from other disorders. Furthermore, Granacher and Berry (2008) advocate for assessment methods based on the identification of non-anatomical or non-physiological presentations of somatic complaints.

Accuracy of Detection Strategies

In order for validity measures to be effective, the cut scores implemented by the test developers have to be accurate enough to identify feigning. Rogers (2018) advised that four

utility estimates be examined when evaluating the effectiveness of detection strategies and validity measures. These include, *sensitivity*, *specificity*, and aspects of *predictive power* [*positive predictive power* (PPP), and *negative predictive power* (NPP)]. When classifying individuals, who malingers, *sensitivity* represents the proportion of individuals, who are malingering, correctly identified by the cut score. For instance, if 9 out of 10 malingerers are identified by a particular cut score, then the sensitivity is .90. In regards to malingering, *specificity* is the number of individuals who are not malingering who had a negative test result divided by all persons without the condition, or otherwise known as the true negative rate (Greve & Bianchini, 2004). A measure has poor sensitivity if a given cut-off produces a large number of false negative errors (i.e., true malingerers going undetected). Poor specificity is characterized by a high number of false positive in which individuals who are responding honestly are categorized as malingering. Sensitivity and specificity are dependent on the classification strategies used by a measure but are independent of base-rates. Predictive power, an index of confidence made that an individual test result is accurate, is reliant on the accuracy of the test and the base rates of the target condition (i.e., malingering) in the population of interest.

Positive predictive power (PPP) is the probability that an individual meeting a particular cut-off will be correctly identified with malingering. For example, if a cut score correctly identifies 9 of 10 malingerers but misclassifies 40 individuals with genuine impairment, the PPP (9/50) is .18. *Negative predictive power* (NPP) is the probability that an individual is not malingering given a negative test result. Accordingly, Rogers (2018) states that in order to establish construct validity of detection strategies, at a minimum, measures should have a PPP of .75 and .50 for sensitivity. Lastly, as stated by Greve and Bianchini (2004), sensitivity and

specificity are vital for accuracy of the measure while PPP and NPP represents a confidence level one can have that the classification of an individual is correct.

Research Designs in Malingering

Simulation research, known-groups comparisons, and differential prevalence designs are the most common research designs used to evaluate malingering (Roger 2008). Simulation, or analogue, designs involve randomly assigning participants into different experimental conditions. They typically involve instructing non-injured individuals to feign deficits in a made-up litigation scenario. Simulation designs are advantageous for examining malingering due to its excellent internal validity due to the standardization methods involved (e.g., standardized instructions, condition, and manipulation checks). Several limitations include limited generalizability to the real-world setting (i.e., weak external validity) because simulators do not face real-world consequences (Rogers, 2018). Therefore, Rogers (2018) recommended a design strategy examining four groups: simulating non-clinical participants; honestly responding non-clinical participants; honestly responding clinical participants; and clinical participants simulating greater impairment than what they actually experience.

Known groups designs involve two discrete and independent stages: the establishment of criterion groups (e.g., bona fide patients and malingerers); and systematic analysis of similarities between criterion groups (Rogers, 2018). Unlike simulation design research, known-groups comparisons have strong external validity as the participants, settings, issues, and incentives overlap with real-world contexts. However, issues remain with this research design regarding the accuracy with establishing criterion groups. Additionally, there is no control over experimental assignment or standardization procedures.

Differential prevalence designs involve researchers inferring differences in group membership based on their inclusion in two different groups. Using litigation and non-litigation as an example, litigants are expected to have higher levels of faking than non-litigants who do not have an incentive to feign. Disadvantages associated with this research design, involve its weak internal validity, as the researcher, again, has no control over experimental assignment or standardization procedures.

Assessment Measures of Malingering

Research and test development on the understanding and detection of malingering has grown since the 1980s. Breting and Sweet (2013) conducted a search with the word malingering and found only one publication in 1985, 18 in 1990, 66 in 2000, and 91 in 2010. For an extended period of time symptom validity testing (SVT) has been synonymous with “effort” testing (Pankratz, 1983). Furthermore, a large percentage of these SVTs were designed to capture cognitive feigning while using the forced-choice testing approach detailed previously. Recently, Larrabee (2012) has advocated for a change in terminology associated with SVTs. In particular, he described performance validity as a better characterization of an individual’s test performance as it is or is not an accurate reflection of their actual level of ability (Bigler, Kaufmann, & Larrabee, 2010; Larrabee, 2012). He contended that it is much more descriptive than previous terms (e.g., effort, symptom validity, or response bias). Therefore, performance validity should assess actual task performance. Furthermore, he further stated that symptom validity should be used to describe the accuracy of symptomatic complaints on self-report measures, such as the MMPI-2-RF. Lastly, embedded validity indicators (EVIs) represent performance validity indicators that are based within standard neuropsychological measures. EVIs are advantageous given that they assess both genuine and feigned deficits, allow for retrospective evaluations of

response validity in previous examinations, expand feigning detection methods employed, and provide some protection against coaching (Rogers, 2018).

Van Dyke and her colleagues (2013) set out to distinguish the characteristics unique and shared between the constructs of cognitive performance, symptom self-report, performance validity, and symptom validity using a confirmatory factor analysis approach. They hypothesized that each of these four concepts would load onto four separate constructs. They used a sample of non-litigating veterans receiving comprehensive neuropsychological evaluations using maximum likelihood estimates and multiple fit indices. Despite hypothesizing a 4-factor model fitting the data, they found that a 3-factor model demonstrated the strongest results suggesting that cognitive performance, performance validity, and self-reported symptoms should be examined separately.

Therefore, due to the growing support for the classification of performance validity and symptom validity, measures of response bias formerly referred to as SVTs will be described as PVTs. Additionally, SVTs will describe measures used to detect the accuracy of symptomatic complaints on self-report measures, such as the MMPI-2-RF, in this study. Therefore, when describing the results related to PVTs and EVIs in this study, the terms performance or performed will be used. Conversely, when describing the results related to SVTs in this study, the terms endorse or endorsement will be used.

Performance Validity Testing (PVT)

Regarding concerns about cognitive abilities and performance, measures have been designed specifically to assess for cognitive response bias (Boone, 2013). Additionally, there is consideration for EVIs derived from standard cognitive tests that have utility. PVTs are effective largely in part to faulty assumptions held by lay individuals regarding brain injuries and their

associated impairments. Furthermore, research has shown that only the most severe brain injuries result in debilitating impairments (Baddeley & Warrington, 1970; Black, 1986; Heaton et al. 1978; Mittenberg et al., 1996; Rawling & Brooks, 1990; Rubinsky & Brandt, 1986; Wiggins & Brandt, 1988).

Forced-choice testing measures present the individual with a large number of items displayed in a multiple-choice format and compares their performance to the likelihood of success based on chance alone (Rogers, 2018). A significant portion of forced-choice measures use the two multiple-choice alternatives. This method asserts that based on chance alone, the individual has a 50% chance of purely guessing the correct response. Scores significantly lower than chance performance suggest that there were intentional efforts to produce incorrect responses. Therefore, a logical conclusion can be made that the below chance probabilities score is a sign of deliberate poor performance. Pankratz and colleagues (1975) implemented the forced choice procedure to assess the validity of self-reported sensory impairment. This study greatly influenced the development of many forced-choice performance validity measures using two alternative forced-choice stimulus presentation, including the Test of Memory Malingering (TOMM; Tombaugh, 1996), Word Memory Test (WMT; Green, Iverson, Allen, 1999), Medical Symptom Validity Test (MSVT; Green, 2004), and the Nonverbal Medical Symptom Validity Test (NV-MSVT; Green, 2010). Modifications to forced-choice testing measures include the addition of more difficult items (Hiscock & Hiscock, 1989) with the Portland Digit Recognition Test (PDRT; Binder, 1990; Binder & Willis, 1991) demonstrating the utility of this technique. This method diminished the need to rely solely on below chance performances as the indicator of feigning.

An important aspect of effective PVTs is high sensitivity and specificity (Boone, 2013). Sensitivity describes the percentage of non-credible individuals identified as such, whereas specificity refers to the proportion of credible individuals identified as such by the test. Akin to a scale, as sensitivity improves specificity decreases and vice versa. Therefore, it is vital to determine the sensitivity and specificity of PVTs for their utility. Cutoff scores are traditionally set to maintain $\geq 90\%$ specificity (Boone, Lu, & Wen, 2005; Boone, Sherman, Palmer, Back, Shamieh et al., 2000; Greve et al., 2008; Kim et al., 2010; Kim et al., 2010) in order to reduce the occurrence of false positive errors.

The majority of PVTs have been validated using simulation studies and known-groups (criterion groups) designs. Simulation studies detail specific instructions to participants, often non-clinical volunteers, to feign in a specific manner. Issues arise with this research design due to the generalizability to real-world non-credible populations, such as age, education, and the stark difference between simulated feigners and active compensation seekers. Known-groups design validation studies, use patients assigned to non-credible groups based on motives to feign, failure on independent PVTs, and/or discrepancies between low cognitive scores and functioning in activities of daily living (ADL). Credible groups are those shown to have zero incentive to feign and fail one or no PVTs (Boone, 2013). Additionally, they should have no diagnoses of amnesic disorder/dementia or low IQ (<70) (Dean, Victor, Boone, & Arnold, 2008; Dean, Victor, Boone, Philpott, & Hess, 2009). Feigning groups should have incentives to feign present and failure of two or more PVTs. Concerns regarding known groups design surface in the accuracy of true criterion groups and the potential for misclassification of individuals based on faulty exclusion/inclusion criteria.

Questions often arise concerning the use of multiple PVTs as they have been suggested to “complicate” test interpretation (Boone, 2013). Various studies have demonstrated that failure on multiple PVTs best discriminate credible and non-credible groups in terms of total classification correct (Chafetz, 2011; Dean et al., 2008; Giger, Merten, Merckelbach, & Oswald, 2010; Larrabee, 2003c; Meyers & Volbrecht, 2003; Sollman, Ranseen, & Berry, 2010; Suhr, Tranel, Wefel, & Barrash, 1997). Research also shows that at times, failure on a single PVT (out of four) is not rare in credible patients in a clinical sample; however, only 5% failed two, and 1.5% failed three, and none failed four (Victor et al., 2009). According to Boone (2013), the administration of several PVTs limits false positive identification as increased PVT failure improves specificity despite not increasing sensitivity. Larrabee (2008) demonstrated that the likelihood of obtaining a false-positive of malingering decreases with each subsequently failed PVT. He reported that chaining of likelihood ratios demonstrates an increase in probability of malingering with increasing numbers of PVTs: .713 to .837 for one failed PVT; .936 to .973 for two failed PVTs; .989 to .995 for three failed PVTs.

Despite the improvement over time in PVTs, individuals can still perform at or below chance without the presence of intentional feigning (Boone, 2008). The presence of genuine cognitive impairment, psychological disorders, the adversarial nature of the evaluation, disengagement from the testing process, and much more, can contribute to poor performances (Rogers & Bender, 2003). Therefore, it remains increasingly important to use a multi-method assessment to improve diagnostic capabilities.

Symptom Validity Testing (SVT)

Many self-report measures are susceptible to response bias due to their inability to detect non-credible response patterns. Thus, some measures useful in clinical settings have limited

utility in forensic contexts (Boone, 2008; Rogers, 2018). As stated previously, several measures have been developed that assess various domains of symptomology and contain numerous validity scales that elucidate the veracity of said complaints.

One of the most widely used self-report psychopathology measures used in both clinical and forensic settings is the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) and its revisions, the MMPI-2 (Butcher et al., 1989; Butcher et al., 2001) and more recently, the MMPI-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008 & 2011). The MMPI-2-RF contains nine validity scales that assess the content and non-content based responses. With regards to this study, the MMPI-2-RF contains five validity scales that measure the over-reporting of symptoms in areas of psychopathology, psychosis, cognitive complaints, and somatic complaints. Additionally, several validity scales were designed specifically for a litigation setting on the MMPI-2-RF. The Symptom Validity Scale-revised (FBS-r; Ben-Porath & Tellegen, 2008) was designed to capture a compensation-seeking response “constellation” of over-reported physical symptoms and exaggerated post-injury emotional distress. An additional validity scale, the Response Bias Scale (RBS; Gervais, Ben-Porath, Wygant, & Green, 2007), empirically predicted failure on cognitive PVTs in a large non-head-injury litigant sample. A detailed review of the literature regarding the MMPI-2-RF over-reporting scales will be discussed later.

Another self-report psychopathology measure similar to the MMPI-2-RF is the Personality Assessment Inventory (PAI; Morey, 1991). It contains seven validity scales of which the Negative Impression (NIM), Malingering Index (MAL; Morey, 1996), and Rogers Discriminant Function (RDF; Rogers et al., 1996) have received the most research attention (Boone, 2013).

The Structured Interview of Reported Symptoms (SIRS; Rogers, 1992) and SIRS-2 (Rogers, Sewell, & Gillard, 2010) are structured clinical interviews designed specifically to assess non-credible psychological symptom reports. Unique to the SIRS and SIRS-2, specific symptoms and severity repeated inquires, as well as general inquiry approaches were implemented. The SIRS has eight primary scales that assess a combination of rarely endorsed items and combinations; generalized lay symptoms; over-reporting of symptoms; and discrepancies between reported and observed symptoms. The SIRS-2 is identical except for a decision tree with two indeterminate classifications.

Some have suggested that the use of multiple PVTs and SVTs raises the probability of false-positives, over and beyond their singular false-positive rates (Berthelson, Mulchan, Odland, Miller, & Mittenberg, 2013). Using a Monte Carlo simulated design, Berthelson and colleagues (2013) found a significant increase in false-positive rate when larger sets of PVTs (each with .15 or .10 per-test false positive rates) were used. Several researchers analyzed data in various samples, the majority not referred for medicolegal evaluations, finding small insignificant correlations between PVTs and SVTs (Davis and Millis, 2014; Larrabee, 2003, 2009, 2014; Schroeder and Marshall, 2011; Victor, Boone, Serpa, Buehler, and Ziegler, 2009). Additionally, methodology/design concerns regarding the Berthelson et al. (2013) study call into question its results. The large body of literature on multiple PVT/SVT failures suggests a high probability of invalid performances. Furthermore, the National Academy of Neuropsychology position statement on symptom validity assessment (Bush et al., 2005) and consensus paper developed by the American Academy of Clinical Neuropsychology (AACN) on effort, response bias and malingering (Heilbronner, Sweet, Morgan, Larrabee, Millis, & the Conference Participants, 2009) advised the use of multiple PVTs and SVTs in clinical and forensic assessment. Additional

researchers have recommended the use of multiple PVTs and SVTs throughout the evaluation (Boone, 2009; Larrabee, 2014, 2015)

Summarily, PVT and SVT failures do not equate to malingering. Various factors must be present, in addition to multiple PVT and SVT failures, such as the context of a substantial external incentive, and a lack of clear evidence of major neurological, psychiatric, or developmental deficits contributing to poor performance meets the general criteria for probable malingering using the MND criteria established by Slick and colleagues.

Embedded Performance Validity Indicators within Standardized Neuropsychological Measures

Malingering can occur in three different fashions in neuropsychological settings (Iverson & Binder, 2000; Larrabee, 2000): (1) fabrication or exaggeration of symptomatic complaints (Larrabee, 1998; Nelson, Sweet, & Demakis, 2006; Wygant, Sellbom, Ben-Porath, Stafford et al. 2007); (2) intentionally poor performance on neuropsychological tests (Binder & Willis, 1991; Mittenberg et al., 1996); and (3) a combination of both (Heaton et al., 1978; Larrabee, 2003). Thus, it is expected that individuals attempting to feign cognitive impairment will indiscriminately perform poorly across performance validity measures and standard neuropsychological measures. Therefore, numerous neuropsychological measures used in standard clinical practice contain embedded “effort” measures that aid in the detection of malingering. The addition of embedded performance validity indicators (EVIs) have the advantage of measuring both response bias and specific cognitive skills without increasing test battery administration time (Boone, 2013).

Heaton et al. (1978) have been credited with first demonstrating that patterns of neuropsychological test performance could discriminate non-injured dissimulators from non-litigating patients with moderate to severe head injuries. Significant differences between the

performance of these two groups were found on the WAIS-III Digit Span (short-term memory), and a number of measures on the Heaton-Reitan Battery (Heaton, Miller, Taylor, & Grant, 2004) such as Category Tests errors, Tactual Performance Test Total Time, Memory and Location, Finger Tapping, and the Hand Dynamometer. These results were replicated by Mittenberg et al. (1996) who demonstrated similar findings. Patterns of performance indicative of poor task engagement have also been reported on individual neuropsychological tests, such as poor recognition memory on the Auditory Visual Learning Test (AVLT; Binder, Villanueva, Howieson, & Moore 1993; Boone, Lu, & Wen, 2005) and poor recognition memory on the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober 1987; Millis, Heilbronner, Sweet, Morgan, Larrabee et al. 1995). Additionally, studies have demonstrated complicated patterns of intentional underperformance including abnormal recall and recognition on the CVLT-II (Delis, Kramer, Kaplan, & Ober, 2000; Wolfe, Millis, Hanks, Fichtenberg et al. 2010). Executive function tests such as the Wisconsin Card Sorting Test (WCST; Bernard, McGrath, & Houston, 1996; Suhr & Boyer, 1999) and the Category Test (Tenhula & Sweet, 1996) have also been used to identify atypical patterns of performance. Regarding the Wechsler Adult Intelligence Scale-IV (WAIS-IV; Wechsler, 2008), the Reliable Digit Span (RDS; Greiffenstein, Baker, & Gola, 1994) was developed as a performance validity score. As part of the WAIS-IV normative data, scores are provided for clinical groups with neurological deficits (e.g., Traumatic Brain Injury, temporal lobectomy, etc.).

MMPI-2-RF and Malingering

The MMPI-2-RF (Ben-Porath & Tellegen, 2008) is the most recent revision of the MMPI-2. It consists of fewer items (i.e., 338 versus 567) and includes revised versions of MMPI-2 validity scales in addition to two new validity scales. It has five over-reporting validity

scales which include Infrequent Responses (F-r), Infrequent Psychopathology Responses (Fp-r), Infrequent Somatic Responses (Fs), Symptom Validity (FBS-r), and the Response Bias Scale (RBS).

Gervais and his colleagues (2007) empirically developed the Response Bias Scale (RBS) specifically to identify persons who over-report cognitive symptoms in forensic neuropsychological or disability assessment settings. RBS was validated using a known-groups design comparing responses of those who failed PVTs (e.g., the Word Memory Test [WMT]) and those who did not. Nelson, Sweet, and Heilbronner (2007) demonstrated its effectiveness in differentiating those with secondary gains versus those without. RBS has been shown to be a better than other MMPI-2 validity scales at predicting memory complaints (Gervais, Ben-Porath, Wygant, and Green, 2008; Wygant et al., 2007). Furthermore, the RBS scores demonstrated zero correlations with objective measures of verbal memory (i.e., CVLT) when performance validity was taken into consideration. Several researchers have evaluated its utility in combination with various performance validity measures. The RBS has been found to predict TOMM failure scores better than other MMPI-2 validity scales (Whitney, Davis, Shepard, and Herman, 2008).

Numerous studies have examined the utility of the MMPI-2-RF across various forensic contexts, including worker's compensation, suspected head-injury, veteran's PTSD compensation, along with criminal and correctional settings.

Sellbom and Bagby (2010) used a simulation design to compare college students who were instructed to feign (coached vs. non-coached) in comparison to psychiatric patients with severe mental disorders. They wanted to determine if coaching impeded the validity scales from discriminating between the two groups. They hypothesized that F-r and Fp-r would have the

greatest discrimination utility. F_{P-r} was superior to all other Validity scales in discriminating over-reporting.

Several studies have examined the ability of the over-reporting scales to discriminate feigned PTSD symptoms in clinical samples including compensation seeking veterans. Goodwin, Sellbom, and Arbisi (2013) used a known groups and simulation design comparing compensation seeking veterans (standard) vs. a simulation group (compensation seeking veterans and mental health professionals asked to feign). They were curious to determine if specific knowledge of PTSD influenced the detection skills of the MMPI-2-RF. $F-r$ and F_{P-r} showed the largest effect sizes across both comparisons. Marion, Sellbom, and Bagby (2011) examined the ability of the MMPI-2-RF over-reporting scales to distinguish feigners of Major Depressive Disorder (MDD), Schizophrenia (SCZ), and Post Traumatic Stress Disorder (PTSD) from true psychiatric patients. Using a simulation design, college students (naïve feigners) and mental health professionals (sophisticated feigners) were asked to feign. F_{P-r} was the best predictor of over-reporting of all disorders.

In a correctional setting, Wall, Wygant, and Gallagher (2015) examined the utility of the MMPI-2-RF validity scales to detect over-reporting. Randomly assigned inmates were asked to feign while another group responded under standard instructions. Additionally, a sample of psychiatric inpatient inmates were recruited to complete the test under standard instructions. They found that $F-r$ best discriminated between feigners and control inmates, while F_{P-r} and F_s discriminated better between feigning and psychiatric inmate groups.

Wygant, Ben-Porath, Arbisi, Berry and colleagues (2009) examined the utility of the somatic/cognitive over-reporting scales (i.e., F_s , $FBS-r$, and RBS) in 3 diverse samples in a combination of simulation and known-group designs. One simulation group consisted of

individuals with a history of head injury compared to those instructed to feign symptoms of a head injury within a disability evaluation context. The second simulation group consisted of medical participants instructed to exaggerate somatic and emotional complaints versus medical patients with no incentive to over-report. The criterion group consisted of a sample of personal injury and disability claimants receiving a psychological evaluation. Large effect sizes were found for F_s in the head injury simulation group; $FBS-r$, $F-r$ and F_s in the medical simulation group; whereas, $FBS-r$ and $F-r$ identified poor task engagement on PVTs.

Sellbom, Wygant, and Bagby (2012) examined F_s utility to detect non-credible somatic complaints between college students asked to feign, patients with bona-fide medical disorders, and patients with somatoform disorders. F_s and F_{p-r} demonstrated the highest differential validity with F_s as the most sensitive to somatic malingering and F_{p-r} was most specific.

Associations between the MMPI-2-RF and malingering criteria have been examined extensively. Wygant, Anderson, Sellbom, Rapier et al. (2011) used the MND (Slick et al., 1999) and Malingered Pain Related Dysfunction criteria (MPRD; Bianchini, Greve, & Glynn, 2005), PVTs [e.g., TOMM, Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Spellacy, 1996), and Letter Memory Test (LMT; Inman, Vickery, Berry, Lamb et al., 1998)], along with symptom validity measures, such as the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) and the MMPI-2-RF to discriminate between malingering and non-malingering. Participants were grouped into 4 malingering categories: incentive only, possible, probable, and definite malingering. Results demonstrated that $F-r$ and RBS discriminated best between incentive only and probable/definite malingering groups.

The SIRS-2, the MMPI-2-RF, and several PVTs were used to investigate feigned mental disorders (FMD) and feigned cognitive impairment (FCI) in a civil forensic sample (Rogers,

Gillard, Berry, & Granacher, 2011). $F-r$ and F_{P-r} produced very large effect sizes discriminating true psychopathology from FMD. Furthermore, FBS-r and RBS were effective in conjunction with other PVTs. Additionally, Taescavage, Wygant, Gervais, and Ben-Porath (2013) examined the over-reporting scales along with MND criteria in non-head injury disability claimants. They hypothesized that FBS-r and RBS would have the greatest association with performance validity tests, whereas RBS would have the greatest association with MND. Individuals were given the MMPI-2-RF and at least 2 PVTs (e.g., TOMM, MSVT, and CARB) and categorized into 4 malingering groups. Results demonstrated that $F-r$ and RBS had the strongest associations with PVT performance while also working best at discriminating between Incentive Only and Probable/Definite MND groups.

Regarding cognitive symptoms, the RBS has been validated in various neuropsychological domains. Particularly, the RBS demonstrated greater incremental validity at assessing over-reporting of memory complaints than its MMPI-2 predecessor (Gervais, Ben-Porath, Wygant, & Sellbom, 2010). The validity scales of the MMPI-2-RF have been examined in head injury neuropsychological evaluations. In a large sample ($n = 501$) of military members completing evaluations for mild traumatic brain injury (mTBI), the authors demonstrated that RBS had the greatest utility in discriminating mTBI military veteran evaluatees who fail no PVTs versus those who fail 3 PVTs (Jones, Ingram, Ben-Porath, 2012). Furthermore, failure on PVTs was associated with a significant increase in all over-reporting scales. McBride, Crighton, Wygant, and Granacher (2013) examined performance validity measure failure in a group of brain-injured individuals with documented intra-cranial brain injury (noted from a CT or fMRI) and head-injury litigants without neuroimaging evidence. They were administered a full

neuropsychological evaluation including 3 PVTs and the MMPI-2-RF. The results demonstrated no significant relationship between the lesion existence, PVT, and SVT measure performance.

Last, a recent meta-analysis of 25 experimental and quasi-experimental studies was conducted by Ingram and Ternes (2016) to investigate the effectiveness of the over-reporting scales to detect feigned symptoms. The results demonstrated that the over-reporting scales demonstrated very large effect sizes, even after accounting for differences in comparison groups. Malingered responding consistently produced a profile of over-reporting validity scales elevated at least 1 SD greater than the study comparison group. Fp-r was the strongest predictor of over-reporting, followed by F-r, Fs, RBS, and then FBS-r. Furthermore, effect sizes accounted for 80% of the variance observed. Moderating factors included respondent diagnosis, referral group, and the comparison group.

Malingering as a Categorical or Continuous Construct

Larrabee (2012) details the need for increased research into subgroups of malingering, specifically, as to whether it occurs dimensionally (along a continuum) or as a dichotomous construct (i.e., malingering vs. honest responding). Taxometric analysis has been a means for examining the latent structure of malingering as it is a useful tool for evaluating the accuracy of classification systems (Meehl & Yonce, 1994; 1996; Ruscio, Haslam, & Ruscio, 2006). This method has been implemented to examine the possibility of taxons of malingering with performance validity and symptom validity measures.

Mean above minus below a cut (MAMBAC; Meehl & Yonce, 1994) and maximum covariance (MAXCOV; Meehl & Yonce, 1996) are two techniques used in determining taxons. MAMBAC works by creating a series of cuts scores that effectively divide a population into two groups on a psychopathology measure. Participants are ordered by their scores on the measure of

interest. Cuts along one indicator are created and differences in scores on a second indicator are examined for cases falling above and below each cut. If the latent structure is of a categorical nature, the presence of an optimal cut score is identified. If no underlying taxon exist, then the plot takes on a disk-shaped curve suggesting a dimensional latent structure (Nathan & Langenbucher, 2003).

MAXCOV works by correlating two indicators with one another across groups by the scores obtained on the indicators (Nathan & Langenbucher, 2003). It requires at least three indicators with at least one being continuous. An example given by Nathan & Langenbucher (2003), in describing two diagnostic criteria for major depressive disorder being correlated with one another in individuals with only, one, two, three, four, or more of the remaining diagnostic criteria for depression. This process is continually repeated using all different combinations of indicators until they all have been correlated with each other.

Strong, Greene, and Schinka (2000) used taxometric analysis to examine the underlying structure of the MMPI-2 F and Fp scales in a psychiatric inpatient and Veterans Affairs (VA) medical center setting. They hypothesized that the endorsement of items on the F and Fp scales would represent an over-reporting group and that overall, there are two distinct taxons; a group answering items on the MMPI-2 in an over-reported manner and a group not responding in this manner. Using two large combined samples, including approximately 50% evaluated for disability claims ($n = 2,030$), mean above minus below a cut (MAMBAC) and maximum covariance (MAXCOV) analyses were performed. Results suggested that F and Fp demonstrated two taxons rather than the degree to which a profile is elevated.

Following the approach of Strong and colleagues (2000), Strong, Glassmire, Frederick, and Greene (2006) examined the latent structure of the Fp scale within a criminal psychiatric

setting. Specifically, these individuals were pretrial defendants referred for various assessment referrals including competency to stand trial. They believed two taxons would be evidenced: those who accurately reported genuine psychopathology and those who over-reported symptoms. Using a similar analytic approach, the results suggested the presence of two distinct groups.

Using the VSVT, Frazier, Youngstrom, Naugle, Haggerty et al. (2007) examined the possibility of a continuum of poor effort/invalid responding or a dichotomy of adequate versus inadequate task engagement. Using a clinical sample of outpatient neuropsychological evaluatees, the results demonstrated evidence of two taxons for cognitive performance pass/failure. Furthermore, those who were grouped in the failure category produced poorer IQ and memory scores relative to those not classified as such.

Conversely, several studies have demonstrated the presence of dimensional latent structure for malingering using a combination of SVTs. Using the six primary scales of the SIRS and several MMPI-2 scales [F, Fp, and Dissimulation (Ds)], in a large sample of criminal and civil evaluations ($n = 1,211$), Walters, Rogers, Berry, Miller and colleagues (2008) examined the possibility of a dimensional latent structure of malingering using the SIRS and a combination of the SIRS and MMPI-2. Both models demonstrated evidence for a dimensional latent structure.

Walters, Berry, Lanyon, and Murphy (2009) examined the possibility of 3 dimensional domains of feigning (psychological, cognitive, and physical) using the Psychological Screening Inventory (PSI; Lanyon, 1970). They used mental health outpatient and a forensic sample while examining the data using a MAMBAC and MAXCOV methodology. Their results suggested the presence of a dimensional latent structure for exaggerated health complaints construct.

Walters, Berry, Rogers, Payne, and Granacher (2009) examined the latent structure of feigned cognitive deficits using the TOMM, LMT, and VSVT in a sample of civil litigant

evaluations. Using MAMBAC, MAXCOV, and latent-mode factor analysis (L-Mode), results showed support for a dimensional latent structure based on all 3 analyses.

Summarily, Larrabee (2012) suggested a need to examine the possibility that the three domains of malingering (psychological, cognitive, physical) fall along a common dimension or several dimensions. This would require the use of a large sample of subjects with incentive to feign in all three domains. Additionally, Meehl (1995) recommends the need to evaluate the prospect of dimensional latent structures using different domains (e.g., self-report, interview based, and performance based) as indicators. Last, Walters et al. (2009) proposed that future studies utilize exploratory or confirmatory factor analysis approaches to elucidate the number of dimensions that underlie the constructs and whether different domains of feigning and their respective detection strategies share common dimensions.

Latent Class Analysis

Latent class analysis (LCA) is a statistical method that can be used to help validate the concerns regarding malingering. LCA works by identifying unobserved (latent) categorical variables (e.g., hypothetical constructs) that account for the covariance between two or more observed (manifest) variables (McCutcheon, 1987; Thomas, Lanyon, Millsap, 2009). This method can be implemented with categorical and/or continuous observed variables, though latent profile analysis is the variation associated with continuous variables.

LCA has been used recently to examine the latent structure of numerous mental disorders related to diagnostic capabilities. It has been implemented in the study of the latent structure of alcoholism (Bucholz, Heath, Reich, Hesselbrock, Krarner et al., 1996), eating disorders (Bulik, Sullivan, & Kendler, 2000; Keel, Fichter, Quadflieg, Bulik et al., 2004), depression (Chen, Eaton, Gallow, & Nestadt, 2000), and social phobia (Kessler, Stein, & Berglund, 1998).

Additionally, it has been used to examine subtypes of Antisocial Personality Disorder (Bucholz, Hesselbrock, Heath, Kramer, & Schuckit, 2000) and analysis of Attention Deficit/Hyperactivity Disorder (ADHD; Rasmussen, Neuman, Heath, Levy et al.2002).

Recent studies have implemented LCA to examine latent constructs on self-report measures (e.g., MMPI-2/MMPI-2-RF). Thomas, Lanyon, and Millsap (2009) demonstrated the utility of LCA to accurately classify individuals using three scales measuring one domain-specific area of misrepresentation: positive impression, exaggeration of virtue, and positive response bias. The three scales used to capture the misrepresentation construct included the Lie (L) Scale of the MMPI-2, Impression Management (IM) scale of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1988 &1991) and the Endorsement of Excessive Virtue (EEV) Scale of the Psychological Screening Inventory (PSI; Lanyon, 1993). With a sample of 209 forensic clients, they examined goodness of fit indices to determine the total number of classes and what structure best fit the data. Latent class analysis was able to classify individuals as either exaggerating virtuous qualities or honest responding.

Again, using the MMPI-2 and latent class analysis, Forbes, Elhai, Miller, and Creamer (2010) examined the typology of posttraumatic stress disorder (PTSD) on the Personality Psychopathology-5 (PSY-5) scales. They proposed three distinct classes of individuals based on the PSY-5 scales. Using a sample of Australian military veterans, they found that a 4-class solution best fit the data rather than the 3 classes they hypothesized. The results showed a simple PTSD class, an externalizing class, and two classes consistent with internalizing traits.

Lastly, Mossman, Wygant, and Gervais (2012) implemented latent class modeling (LCM) to determine if it had applications for generating inferences about the accuracy of PVTs used to capture potential malingering in real-world data. Their rationale for this study involved

the lack of a “gold standard” for capturing malingering given that most PVT measures were developed using simulation and/or known group designs. They posited that LCM provides accuracy estimates for PVTs without the disadvantages of the simulation or known group design. They used the TOMM, WMT, and Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997) and a sample of forensic evaluations estimated the accuracy of these PVTs. They found that the PVT measures performed better than chance at detecting feigned cognitive impairment but that the WMT was far superior to the CARB or TOMM. Their findings demonstrated that PVT scores were able to estimate accuracy parameters for each test in the absence of “an infallible criterion for ascertaining whether the subjects were performing at less than their true ability” (Mossman et al., 2012).

The Present Study

The current study sets to investigate the possibility of subgroups of malingering. Previous studies have found support for either a categorical or continuous latent structure regarding malingering; however, none of these studies evaluated more than one measure or test domain in their analyses. Many of the measures included either scales from the MMPI-2 and SIRS (self-report inventories), a single PVT, or three PVTs from the same test domain: verbal memory recognition paradigm with forced-choice procedures (i.e., VSVT, TOMM, LMT). Furthermore, none of these studies have implemented the latent class modeling approach to investigate subgroups of malingering. Numerous studies and several diagnostic malingering criteria (Slick et al., 1999; Bianchini et al., 2005) suggest that malingerers are grouped into discrete categories.

CHAPTER II

METHOD

The current study used a large civil litigation sample receiving three PVTs, a SVT, and an EVI as part of a comprehensive forensic psychological evaluation. In this study, I determined what distinct group profiles emerge with the use of a diverse set of validity measures. Specifically, measures that capture performance, symptom and embedded performance validity were used. Measures from these domains were implemented given that numerous studies have shown that malingering manifests in different manners (e.g., symptom exaggeration vs. marked inconsistencies on performance measures; Rogers, 2018) and should be assessed with different measures at multiple points in time (Boone, 2007) Additionally, examination of these groups was made to determine if they differ with respect to the validity measures. Exploratory latent class analysis was implemented to identify the classes underlying the latent structure of malingering.

Participants

The data came from an archival dataset of 4,631 civil litigation claimants evaluated at a private forensic psychological practice, within the context of a personal injury or disability evaluation. Individuals were administered psychological, neuropsychological, and intellectual measures in a standardized order (e.g., standard-fixed battery) that extended typically over a day and a half of testing. As part of their evaluation consent, the informed consent stated that the test battery contained validity checks and that individuals were asked to perform and respond to all items to the best of their abilities. This dataset was part of an ongoing data collection project by Dr. Roger Gervais which began in 1996. Subsets of the dataset have been used in other studies (Armistead-Jehle & Gervais, 2011; Armistead-Jehle, Gervais, & Green, 2012a, 2012b; Armistead-Jehle, Green, Gervais, & Hungerford, 2015; Demakis, Gervais, Rohling, 2008;

Gervais, Ben-Porath, Wygant, 2009; Gervais, Ben-Porath, Wygant, Green, 2007, 2008; Gervais, Ben-Porath, Wygant, & Sellbom, 2010; Gervais, Green, Allen, & Iverson, 2001; Gervais Rohling, Green, & Ford, 2004; Gervais et al., 2001; Gervais, Wygant, Sellbom, & Ben-Porath, 2011; Green, Rohling, Iverson, Gervais, 2009; Greiffenstein et al., 2010; Iverson, Green, Gervais, 1999; Lee, Graham, Sellbom, & Gervais, 2012; Mossman et al., 2012; Richman, Green, Gervais, Flaro et al., 2006; Sellbom, Lee, Ben-Porath, Arbisi et al., 2012; Tarescavage et al., 2013; Wiggins, Wygant, Hoelzle, & Gervais, 2012; Wygant, Sellbom, Gervais, Ben-Porath et al., 2010). In that sense, the present analyses were a secondary data analysis; however, new cases were added to the dataset on an ongoing basis. Therefore, the analyses included a subset of previously unused cases.

Cases were initially selected from the larger data set if they were given three performance validity tests (PVTs): the Word Memory Test (WMT), the Medical Symptom Validity Test (MSVT), and the Nonverbal Medical Symptom Validity Test (NV-MSVT). This left 1,158 cases for the latent class analysis. Additionally, all participants were given the Reliable Digit Span (RDS) of the Wechsler Adult Intelligence Scale –IV (WAIS-IV), and the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF). The sample was predominately male with 62.6 % and female subjects making up the remaining 37.4% of the sample. The mean age of the patients was 43.1 years old ($SD = 11.2$). The sample had a mean education of 11.8 years ($SD = 2.3$). MMPI-2-RF protocols were excluded from the analysis if they demonstrated evidence of non-content based responding [i.e., Variable Response Inconsistency-Revised (VRIN-r) or True Response Inconsistency (TRIN-r) ≥ 80] and/or by not answering more than 15 items on the MMPI-2-RF (Can Not Say > 14 ; CNS). Based on these criteria, 32 cases were removed for a total sample of 1,126 cases. Assessment type consisted largely of psychological evaluations

(87.9%), followed by vocational (9.1%), with the smallest percentage of other referrals constituting the rest of the sample. Over 80% of the sample was involved in a work-place accident, while a smaller percentage (12.6%) was involved in a motor vehicle accident. See Tables 2 and 3 for the full list of assessment and accident types. With regards to referral type, 82.8% were referred directly from the Worker's Compensation Board (WCB), 10.7% were legal referrals. The smallest percentage was comprised of WCB/Legal, other, private, or insurance referrals. The full list of referral types is listed in Table 4.

Table 2.

Assessment Type.

	N (%)
Pain Management/Counseling	2 (0.2)
Psychological	990 (87.9)
Vocational	103 (9.1)
MLD	30 (2.7)
Other	1 (.1)

Table 3.

Accident Type.

	<i>N (%)</i>
Work Accident	919 (81.6)
Motor Vehicle Accident	142 (12.6)
Neurological	1 (0.1)
No Accident	29 (2.6)
Other	35 (3.1)

Table 4.

Referral Type.

	<i>N (%)</i>
Worker's Compensation Board (WCB)	932 (82.8)
Legal	121 (10.7)
WCB/Legal	1 (.1)
Other	16 (1.4)
Private	3 (0.3)
Insurance	53 (4.7)

Diagnostically, a significant portion of the cases were diagnosed using the Diagnostic and Statistical Manual-IV (APA, 1994) or DSM-IV-TR (2000) criteria and determined by the evaluating psychologist at the time of assessment, following a detailed clinical interview and review of psychological test data and accompanying medical or other third-party documentation. All participants were administered an extensive psychological test battery. Diagnoses largely consisted of chronic pain (19.0%), depressive disorders (35.3%), and anxiety-related disorders (39.1%). Table 5 contains the full list of diagnoses.

Table 5.

Diagnoses.

	<i>N</i> (%)
Chronic Pain	214 (19.0)
Depressive Disorders	398 (35.3)
Anxiety Disorders	440 (39.1)
Bipolar Type Disorders	9 (0.8)
Orthopedic	23 (2.0)
Chronic Fatigue	1 (0.1)
Neurological Disorders	1 (0.1)
Mild TBI	8 (0.7)
Moderate/Severe TBI	6 (0.5)
Psychotic Disorders	1 (0.1)
Cluster A Personality Traits	2 (0.2)
Cluster B Personality Traits	3 (0.3)
Cluster C Personality Traits	2 (0.2)
Other	18 (1.6)

Instruments and Measures

Several instruments and measures were used in the present study. Measures were categorized as either performance validity test (PVT), symptom validity test (SVT), or embedded performance validity indicators (EVI).

Performance Validity Tests (PVT)

Word Memory Test (WMT). The WMT (Green, Allen, & Astner, 2003) is a computer-administered verbal memory test containing multiple subtests of varying difficulty designed to assess task engagement. The WMT involves the presentation of 20 simple word pairs at a rate of one-word pair per two seconds in two learning trials. Additionally, the words are characterized as either “easy” (having an obvious semantic relationship) or “difficult” (less obvious semantic relationship). After the two learning trials, the examinee completes an immediate 40-item forced-choice recognition test (Immediate Recognition; IR). This involves the presentation of each of the words from the previous 40 with an additional incorrect word that was not presented during the learning trials. The patient is presented with a similar setup following a 30-minute delay (Delayed Recognition; DR). Each trial is scored based on the individual selecting the correct word from a word-pair containing a distractor word. After being presented both trials, a consistency of responses score is calculated (Consistency; CNS). Children and adults with documented neurological impairments perform exceedingly well, demonstrating that the WMT is sensitive enough to discriminate poor task engagement from true neurological deficits. It has also been shown that the WMT is highly sensitive to task engagement and demographic or psychosocial variables (e.g., age, gender, race/ethnicity, social-economic status), intelligence, and psychopathology (Green, Allen, & Astner, 2004). Green (2003) suggests a cutoff of less than or equal to 33/40 (82.5%) on the IR, DR, or CNS as indicators of poor task engagement. Tan,

Slick, Strauss, and Hultsch (2002) demonstrated a sensitivity and specificity of 100% for individuals simulating feigned cognitive impairments and those instructed to respond with no cognitive impairment. Additionally, 30% of those instructed to feign scored below chance.

Medical Symptom Validity Test (MSVT). The MSVT (Green, 2004) is a computer administered verbal memory screening test developed to capture poor task engagement. While similar to the WMT, the MSVT was designed to be shorter and easier. It requires an administration time of only 5 minutes in most cases (Green, 2003; Merten, Green, Henry, Blaskewitz et al., 2005; Richman et al., 2006) and contains only 10 word pairs instead of 20. Furthermore, the delay period between subtests is 10-minutes instead of 30 minutes, further decreasing testing time. Each word pair represents one concept instead of two concepts per pair as seen in the WMT (e.g., soccer-ball vs. tree-lake). After being presented twice with a list of word pairs, the individual performs an Immediate Recognition task (IR) and a Delayed Recognition trial 10 minutes later. In both IR and DR trials, the patient is required to choose the word from the original list from 20 new word pairs that includes a foil word (e.g., “ballpoint” from “ballpoint-iron”). A consistency of responses score (CNS) is calculated from the IR to DR, where the patient obtains one point if they correctly choose the correct response or if the incorrect response was chosen both times. The patient obtains a score of zero for any inconsistencies noted for a given item (i.e., scoring incorrectly on one trial and correctly on the other). The patient passes the MSVT if IR, DR, and CNS scores are all above 85%. Failure of the MSVT is denoted by scores at or below 85% on at least IR, DR, or CNS. Green (2004) demonstrated, in the test manual, that children diagnosed with intellectual disabilities, or psychiatric, or neurological conditions were effectively able to pass the MSVT at the 85% cut-off.

Nonverbal Medical Symptom Validity Test (NV-MSVT). The NV-MSVT (Green, 2008) is a computer administered nonverbal memory screening test containing four primary task engagement subtests. Ten artist-drawn colored images, each containing a pair of items strongly associated together are presented on the computer. The items are presented twice on the screen, at a rate of one pair every four seconds. After the list presentation, the patient completes a 20-item-forced choice-recognition trial (Immediate Recognition; IR) that includes a target item from the original list and a foil item not previously seen. After a ten-minute delay, again, a target item is paired with a new foil item and the patient completes a new 20-item forced-choice recognition trial (Delayed Recognition; DR). The Delayed Recognition portion of the NV-MSVT also contains two additional subtests, the Delayed Recognition-Variations (DRV) and the Delayed Recognition-Archetypes (DRA). In the DRV subtest, an original target item (e.g., milk cow) is presented with an almost identical foil but has a slight variation in detail (e.g., milk cow missing a spot). In the DRA subtest, each foil item, that was presented in the IR trial, is paired with a single archetypal image (e.g., snake, a bat, a lion roaring, etc.) and the patient is required to select the foil item previously seen on the computer screen. Lastly, on the Paired Associate (PA) subtest, the patient is shown a major portion of a target item pair (e.g., horse) and is required to determine the missing part of the pair (i.e., cart). Again, similar to the WMT and MSVT, children and adults with intellectual disability or neurological disorders effectively “passed” the validity measure at the 90% cut-off. Additionally, Green (2004) demonstrated that patients with various cortical degenerative processes (e.g., Alzheimer’s disease) successfully passed the NV-MSVT, despite demonstrating significant impairment on measures of executive functioning. Therefore, passing on the NV-MSVT is based on the mean of IR, DR, CNS, DRA, DRV, and PA being above 90% and the mean of DR, CNS, DRA and DRV being at or above 88%. Cut-off for

failure is based on either the mean of IR, DR, CNS, DRA, DRV, and PA being at or below 90% or the mean of DR, CNS, DRA, and DRV being below 88%.

Symptom Validity Test (SVT)

Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF). The MMPI-2-RF (Ben-Porath & Tellegen, 2008) is a self-report personality inventory, comprised of 338 items related to symptoms, beliefs, and attitudes associated with personality and psychopathology. The over-reporting validity scales were examined in the latent class analysis. These scales include Infrequent Responses-Revised (F-r), Infrequent Psychopathology Responses-Revised (Fp-r), Infrequent Somatic Responses (Fs), Symptom Validity Scale-Revised (FBS-r), and Response Bias Scale (RBS). Infrequent Responses-Revised (F-r) is comprised of 32 items rarely endorsed by members of the normative sample. Infrequent Psychopathology Responses-Revised (Fp-r) contains 21 items rarely endorsed by individuals with severe psychopathology and members of the normative sample. Fs contains 16 items with somatic content rarely endorsed by medical and chronic pain patients receiving medical treatment and members of the normative sample. FBS-r contains 30 items associated with non-credible presentation of somatic and cognitive symptoms. Finally, Response Bias Scale (RBS) contains 28 items associated with poor performance on cognitive response bias measures that capture unusual combinations of cognitive and memory complaints. The MMPI-2-RF manual (Ben-Porath & Tellegen, 2008/2011) provide extensive data regarding the psychometric characteristics of the validity scales in a wide variety of samples, both clinical and forensic.

Embedded Performance Validity Indicator (EVI)

Reliable Digit Span (RDS). The RDS (Greiffenstein, Baker, & Gola, 1994) is a performance validity indicator derived from the Digit Span (DS) subtest of the Wechsler Adult

Intelligence Scale–Fourth Edition (WAIS-IV; Wechsler, 2008a; 2008b). The Digit Span test is a measure of auditory attentional capacity that requires an individual to repeat a series of digits forward, backward, and in ascending order. RDS is calculated by summing the longest forward and backward digit sequences for which both trials were completed without error. Research has shown that participants with genuine neurocognitive difficulties performed well on RDS (Greiffenstein et al., 1994b). They found that an RDS score of less than 7 was suggestive of negative response bias. Sensitivity of .57 and high specificity of .93 were found for discriminating probable malingerers from patients with documented head injuries.

Classification of Validity Measures for Latent Class Analysis

Three stand-alone performance validity tests (i.e., the WMT, MSVT, and NV-MSVT), one embedded performance validity indicator (i.e., RDS), and five symptom validity scales (i.e., F-r, Fp-r, Fs, FBS-r, and RBS) were used in the present study, for a total of nine latent variable indicators. Following the cutting scores and decision rules detailed in each performance validity measure's manual, performance on each measure was rated a score between 0, 1, and 2. A score of 0 represented a passing performance on the measure and a score of 1 was given when a patient's score fell at or below the test's cutoff but not below chance. A score of 2 was assigned to any performance on the test that fell at or below chance performance, as noted by the test manuals. Again, when describing the results of the PVTs, the terms *perform* and/or *performance* were used. This is also a similar classification approach implemented by Wygant et al. (2011) who based malingering criteria (Bianchini et al., 2005; Slick et al., 1999) Specific scoring criteria for each performance validity test is illustrated in Table 6.

Table 6.

Scoring Criteria for Performance Validity Test (PVT) Performance.

Test	PVT Performance		
	Pass (0)	Below Cut-off Failure (1)	Below Chance Failure (2)
WMT	>33 on IR, DR, and CNS (>82.5%)	≤33 on IR, or DR, or CNS	≤15 on IR, or DR, or CNS (37.5%)
MSVT	>17 on IR, DR, and CNS (>85%)	≤17 on IR, or DR, or CNS	≤6 on IR, or DR, or CNS (30%)
NV-MSVT	Mean of IR, DR, CNS, DRA, DRV, PA is > 90% And Mean of DR, CNS, DRA, & DRV ≥ 88%	Mean of IR, DR, CNS, DRA, DRV, PA is ≤ 90% Or Mean of DR, CNS, DRA, & DRV < 88%	IR, or DR, or CNS < 30% Or DRA, or DRV, or PA < 20%

Note. WMT = Word Memory Test. MSVT = Medical Symptom Validity Test. NV-MSVT = Nonverbal Medical Symptom Validity Test. IR = Immediate Recognition. DR = Delayed Recognition. CNS = Consistency. DRV = Delayed Recognition-Variations. DRA = Delayed Recognition-Archetypes. PA = Paired Associates.

Based on the cutting scores and decisions rules for the embedded performance validity indicator, performance was rated a score between 0 and 1. A score of 0 represented a passing performance on the measure and a score of 1 represented below cutoff failure. Similar to the PVTs, the terms *perform* and/or *performance* were used to describe the results. Specific scoring criteria for the embedded performance validity indicator is illustrated in Table 7.

Table 7.

Scoring Criteria for Embedded Performance Validity Indicator (EVI).

Test	EDI Performance	
	Pass (0)	Fail (1)
RDS	> 6	≤ 6

Note. RDS = Reliable Digit Span.

Following the cut score ranges and decision rules detailed in the MMPI-2-RF manual, performance on F-r was rated with a score between 0 and 4, Fp-r was rated on a score between 0 and 3, while, Fs, FBS-r, and RBS were rated between 0 and 2. 0 represents a passing score (i.e., valid responding for all scales. A *T* score of 120 on F-r represents invalid responding (i.e.,

definite response bias). A T score of greater than or equal to 100 on Fp-r, Fs, FBS-r, or RBS represents invalid responding (i.e., definite response bias). When describing the results related to SVT performance, the term *endorse* was used. Scoring criteria for the validity scales are illustrated in Table 8.

Table 8.

Scoring Criteria for Symptom Validity Test (SVT) Endorsement on the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF).

Validity Scales	SVT Performance				
	Pass (0)	(1)	(2)	(3)	(4)
F-r	$T < 79$	$T = 79 - 89$	$T = 90 - 99$	$T = 100 - 119$	$T = 120$
Fp-r	$T < 70$	$T = 70 - 79$	$T = 80 - 99$	$T \geq 100$	-
Fs	$T < 80$	$T = 80 - 99$	$T \geq 100$	-	-
FBS-r	$T < 80$	$T = 80 - 99$	$T \geq 100$	-	-
RBS	$T < 80$	$T = 80 - 99$	$T \geq 100$	-	-

Note. MMPI2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; MMPI-2-RF Validity Scales: F-r = Infrequent Responses-Revised. Fp-r = Infrequent Psychopathology Responses-Revised. Fs = Infrequent Somatic Responses. FBS-r = Symptom Validity Scale-Revised. RBS = Response Bias Scale.

Analyses

Latent class analysis (LCA; McCutcheon, 1987) is a mixture modeling technique developed to reveal unobserved heterogeneity within a given population and to define substantively meaningful groups of people that are similar in their responses to measured variables in a cross-sectional manner (Muthen, 2004). LCA models identify categorical latent class constructs measured by a number of observed response variables (i.e., validity indicators). Additionally, LCA works to place individuals into classes based upon the observed items (i.e., person-centered) and highlight items that best capture each class.

Latent class analysis can also be implemented from an exploratory or confirmatory standpoint. Exploratory latent class analysis (ELCA) does not contain a strong a priori hypothesis regarding the number or nature of the latent classes underlying the data (Hojtink, 2001). Advantages regarding an exploratory LCA involve the ability to fit several proposed models to the data with differing numbers of latent classes. Additionally, this allows one to compare the resulting fit indices to determine which best corresponds to the observed data (Finch & Bronk, 2011). Furthermore, exploratory LCA works under the assumption that a validated theory regarding the types of latent groups is present (Laudy, Boom, & Hoijtink, 2005). Confirmatory LCA (CLCA) contrasts ELCA by allowing for the formulation of specific hypotheses regarding the characteristics of the latent classes in the data. CLCA hypotheses are expressed as a set of parameter constraints for an estimated LCA model (Croon, 1990).

When considering an LCA model, *item and class probability parameters* are to be examined. When examining categorical outcomes (e.g., malingering or not malingering), *item parameters* correspond to the conditional item probabilities and represent *measurement parameters*. Alternatively, this refers to the probability individuals may respond to a given item

within a specific class. *Class probability parameters* refer to the prevalence rates or size of each class (i.e., *structural parameters*). Stated differently, this refers to the proportion of individuals belonging to a specific class. For LCA models using continuous outcomes, otherwise known as latent profile analysis, item parameters are class specific item means and variances. Furthermore, the conditional independence assumed for this model suggests that the correlation among the categorical and continuous outcomes is explained by the latent class variable, malingering (i.e., local independence).

Several factors were considered to determine the number of classes needed to define the construct of malingering, otherwise known as class enumeration or model fit (Nylund, Asparohov, Muthen, 2007). Statistical information criteria (IC) such as Akaike Information Criterion (AIC; Akaike, 1987), Bayesian Information Criterion (BIC; Schwartz, 1978), the sample size adjusted BIC (aBIC), and a credible theory are necessary in the determination of classes. Nylund et al. (2007) conducted a monte carlo simulation study to evaluate numerous model fit indices and to give insight as to the most useful indices to use for determining the number of classes. They found that the sample size adjusted Bayesian information criterion (aBIC) was superior, across all models and all sample sizes, to the Akaike information criterion (AIC), the consistent AIC (CAIC), and the standard BIC. The aBIC utilizes the likelihood ratio statistic and applies a penalty for an increased number of model parameters. Furthermore, it is useful for comparing the fit of multiple models, with lower values indicating a relatively better model fit (Finch and Bronk, 2011).

In addition to investigating information criterion, Nylund et al. (2007) examined the performance of three hypothesis testing methods for determining model fit: the chi-square-based likelihood ratio test (LRT), the Lo-Mendell-Rubin (LMR) LRT test, and the bootstrap likelihood

ratio test (BLRT). These LRTs are used to test the null hypothesis (k) to the less restrictive model ($k - 1$). Therefore, a statistically significant p -value suggests that the model fits the data better than the model with one less class ($k - 1$). Nylund and her colleagues (2007) also demonstrated that the chi-square difference test (*naïve chi-square*) is ineffective for class enumeration in LCA. Their results suggested that the Lo-Mendell-Rubin (LMR) and the bootstrap likelihood ratio test (BLRT) are most appropriate for comparing mixture models with differing number of classes. Additionally, the BLRT performs best in LCA models with continuous outcomes. Summarily, Nylund et al. (2007) suggests the BLRT is the better statistical tool of all indices and test, though have disadvantages that necessitate consideration for the LMR.

Therefore, following the recommendations of Nylund and colleagues (2007), an iterative modeling approach and commonly accepted fit criteria were used. This dissertation employed the following information criteria to measure model fit: the Akaike Information Criterion (AIC); the Bayesian Information Criterion (BIC); and the sample size adjusted BIC (Sclove, 1987). It also implemented several likelihood ratio tests such as, the Bootstrap Likelihood Ratio Test (BLRT; McLachlan & Peel, 2000) and the Lo-Mendell-Rubin LRT.

Regarding the theoretical understanding of malingering, several diagnostic criteria have been developed that capture malingering subgroups. Malingered Neurocognitive Dysfunction (MND; Slick et al., 1999) is a set of malingering criteria that was used in this study. Table 1 shows a description of the criteria for the determination of MND.

CHAPTER III

RESULTS

Descriptive Statistics

Table 9 provides a breakdown of the performance rates for each level of the PVT measures based on the evaluatees' scores. Table 10 provides the percentage of individuals who performed at each level of the EVI. Lastly, Table 11 provides the percentage of endorsement rates at each level of the SVT scales.

Table 9.

Total and Percentages of PVT Performance by Levels.

	Pass	Below Cut off	Below Chance
WMT	776 (68.9%)	344 (30.6%)	6 (0.5%)
MSVT	924 (82.1%)	197 (17.5%)	5 (0.4%)
NV-MSVT	935 (83%)	179 (15.9%)	12 (1.1%)

Note. WMT = Word Memory Test. MSVT = Medical Symptom Validity Test. NV-MSVT = Nonverbal Medical Symptom Validity Test.

Table 10.

Total and Percentages of EVI Performance by Levels.

	Pass	Below Cutoff
RDS	1041 (92.5%)	85 (7.5%)

Note. RDS = Reliable Digit Span.

Table 11.

Total and Percentages of SVT Endorsement by Levels.

Scale					
F-r	<i>T</i> < 79 (0)	<i>T</i> = 79 – 89 (1)	<i>T</i> = 90 – 99 (2)	<i>T</i> = 100 – 119 (3)	<i>T</i> = 120 (4)
	584 (51.9%)	184 (16.3%)	101 (9%)	149 (13.2%)	108 (9.6%)
Fp-r	<i>T</i> < 70 (0)	<i>T</i> = 70 – 79 (1)	<i>T</i> = 80 – 99 (2)	<i>T</i> ≥ 100 (3)	-
	943 (83.7%)	84 (7.5%)	75 (6.7%)	24 (2.1%)	
Fs	<i>T</i> < 80 (0)	<i>T</i> = 80 – 99 (1)	<i>T</i> ≥ 100 (2)	-	-
	738 (65.5%)	262 (23.3%)	126 (11.2%)		
FBS-r	<i>T</i> < 80 (0)	<i>T</i> = 80 – 99 (1)	<i>T</i> ≥ 100 (2)	-	-
	586 (52%)	481 (42.7%)	59 (5.2%)		
RBS	<i>T</i> < 80 (0)	<i>T</i> = 80 – 99 (1)	<i>T</i> ≥ 100 (2)	-	-
	589 (52.3%)	359 (31.9%)	178 (15.8%)		

Note. MMPI2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; MMPI-2-RF Validity Scales: F-r = Infrequent Responses-Revised. Fp-r = Infrequent Psychopathology Responses-Revised. Fs = Infrequent Somatic Responses. FBS-r = Symptom Validity Scale-Revised. RBS = Response Bias Scale.

The presentation of the results of the latent class analysis is divided into two sections. The first section provides the evaluative information for the models tested to determine the most appropriate number of classes. This included details about the values for the information criteria (e.g., AIC and BIC), model fit indices (e.g., LMR), and entropy for 1 – 7 class models. Entropy is an indicator of latent class identification (Asparouhv & Muthen, 2018). Entropy values range from 0 to 1 and values approaching 1 indicate clearer separation of classes (i.e., good classification); however, entropy is not a good indicator for a well-fitting model and should not be used in determining the number of classes. The models were run with Mplus Version 7.4 using full information maximum likelihood estimation (Muthen & Muthen, 2015). The second section included an evaluation of the structural and measurement parameter estimates. This included examination of class size, probability of latent class membership, the conditional probabilities (i.e., endorsement of each variable and the categories in each variable). Lastly, the number of criteria for inclusion in each class met were detailed.

Model fit statistics suggested a four or five class model solution since not all the fit statistics pointed at one particular model (see Table 12). After examining the item probabilities for both the four class and five class mode, the five-class solution was selected because it made the most substantive sense. The aBIC and LMR which are used to assess LCA models, suggested a five-class model while the BIC suggested a four-class model (Nylund et al., 2007). While the BIC, aBIC, and LMR values were the lowest for the 5-class model, the AIC stopped decreasing in magnitude sharply, and appeared to tail off between four and five classes. The BLRT was not useful as it was significant for each class model. This is particularly common with real world data as the BLRT appears to work better with simulation data (Muthen & Muthen, 2009).

Table 12.

Model Fit Information for 1 - 7 Class LCA Models.

# of Classes	1	2	3	4	5	6	7
LL	-7391.81	-6408.39	-6231.023	-6073.90	-6005.15	-5971.84	-5946.92
# of parameters	20	41	62	83	104	125	146
BIC	14924.14	13104.87	12897.69	12731.00	12741.05	12821.98	12919.67
ABIC	14860.61	12974.64	12700.76	12467.37	12410.72	12424.94	12455.95
AIC	14823.61	12898.78	12586.05	12313.81	12218.30	12193.68	12185.83
LMR p-value	NA	< 0.001	0.131	< 0.001	0.006	0.439	0.891
BLRT p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Entropy	NA	0.87	0.85	0.84	0.84	0.84	0.86

Note: BIC = Bayesian Information Criteria; ABIC = Adjusted Bayesian Information Criteria;

AIC = Akaike Information Criteria; LMR = Lo-Rubin-Mendell Likelihood Ratio Test; BLRT =

Bootstrap LMR. Bolded values in the tables denotes the model that was preferred by the given fit

index.

Table 13 contains the final class counts and proportions for the latent classes based on their most likely latent class membership. This table describes the individuals with their highest probability for being in that particular class. Tables 14 and 15 show the average latent class probabilities for most likely latent class membership (row) by latent class (column) and classification probabilities for the most likely latent class membership by latent class, respectively. Table 14 represents the classification probability for the individuals in that particular class and their average probability for being in each class.

Table 13.

Final Class Counts and Proportions for the Latent Classes Based on Their Most Likely Latent Class Membership.

Latent Class	<i>N</i>	%
1	83	7.4
2	102	9.1
3	162	14.4
4	264	23.4
5	515	45.7

Table 14.
Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column).

	1	2	3	4	5
1	0.878	0.042	0.078	0.002	0.000
2	0.027	0.890	0.016	0.040	0.026
3	0.018	0.002	0.901	0.079	0.000
4	0.000	0.024	0.054	0.847	0.075
5	0.000	0.012	0.000	0.051	0.937

Table 15

Classification Probabilities for the Most Likely Latent Class Membership by Latent Class.

	1	2	3	4	5
1	0.927	0.035	0.038	0.000	0.000
2	0.000	0.849	0.004	0.058	0.057
3	0.038	0.010	0.867	0.085	0.000
4	0.001	0.015	0.048	0.837	0.099
5	0.000	0.005	0.000	0.039	0.956

Table 16 includes the conditional item probabilities for each class. To visually understand the latent classes, Figure 1 presents the conditional item probability profile plots with the nine validity indicators for each latent class. Item profile plots are used to understand and label the latent classes. The x-axis represents each level of the nine validity indicators. The y-axis is the probability of individuals in a given class endorsing or performing at each level of the validity indicators. To interpret and label the latent classes, the probabilities of endorsing or performing on a particular validity indicator at a particular level were considered. Consideration was also given to differences and similarities between the other classes.

The Latent Classes

Class 1, denoted by the square symbol in Figure 1, was labeled the *Definite Malingering* class. This class included 7.4% of the total sample. Individuals in this class demonstrated a global pattern of performance failure and symptom over-endorsement. The median number of criteria that were at least failed below cut-off and/or indicated possible over-reporting was seven but ranged from five to nine. On the two of the PVTs, over 95% of the individuals in this class at least scored below cutoff demonstrating a very small percentage of passing (i.e., valid responding). Additionally, individuals in this class had the highest percentage of below chance performance on all three PVTs. The largest percentage of RDS failure was seen in the class, as well, at 0.311. Lastly, three out of five symptom validity scales were endorsed at the highest level (i.e., invalid responding on F-r, FBS-r, and RBS) as compared to all other classes.

Class 2, denoted by the triangle symbol, comprised 9.1% of the total sample. It was labeled the *Probable Neurocognitive Malingering* class. For most of the individuals in this class, six criteria were at least failed below cut-off and/or indicated possible over-reporting, though the total number of criteria ranged from three to eight. Over 65% of these individuals scored below

cut-off on the PVTs with 92% failing the WMT. Different from the *Definite Malingering* class, individuals in class 2 had minimal below chance performance. This class also had a large portion of individuals fail the embedded validity indicator. Again, different from the first class, class 2 largely passed most of the SVT scales (i.e., scored 0) as compared to the first and third classes.

Class 3, denoted by the diamond symbol, comprised 14.4% of the total sample. This class was labeled *Probable Symptom Malingering*. Most individuals in this class performed below cut-off and/or endorsed at least possible over-reporting on five criteria but there was a range from two through seven criteria. Individuals in this class passed the PVTs at a higher level than either class 1 or 2, while only 1.4% failed below chance on the NV-MSVT. Individuals in this class also largely passed the RDS as well. With regards to SVT scales, two out of five SVT scales in this class were endorsed at the highest level indicating invalid responding (i.e., F-r and Fs). Additionally, the other three SVT scales in this class had the second highest level of invalid responding (i.e., Fp-r, FBS-r, and RBS).

Class 4, denoted by the circle symbol, comprised 23.4% of the total sample. This class was labeled the *Possible Malingering* class. Most individuals in this class performed at or below cut-off and/or indicated possible over-reporting on three criteria (ranged from one to five criteria). Individuals in this class passed the PVT and EVI measures at a level greater than 85%. Additionally, there was no evidence of below chance performance on the PVT measures. Regarding the SVT scales, there was still evidence of over-reporting and some evidence of invalid responding, albeit, it was lower than that in Classes 1, 2, or 3.

Class 5, denoted by the plus symbol was comprised of the largest percentage of the total sample at 45.7%. This class was labeled the *Valid Responders*¹ class. Most individuals in this

¹ *Valid Responders* class refers to profiles that are within normal limits on the validity measures; however, this does not insure that some individuals in this class may have engaged in response bias.

class endorsed no criteria, meaning that most did not perform at or below cut off on the PVTs and EVI while also demonstrating minimal evidence of over-reporting; however, the range was zero to three criteria. Two out of three PVTs had a failure rate of less than 5% percent. Except for a very small percentage performing below chance (i.e., .3% below chance performance on NV-MSVT), none of the other PVT measures were endorsed at the below chance level.

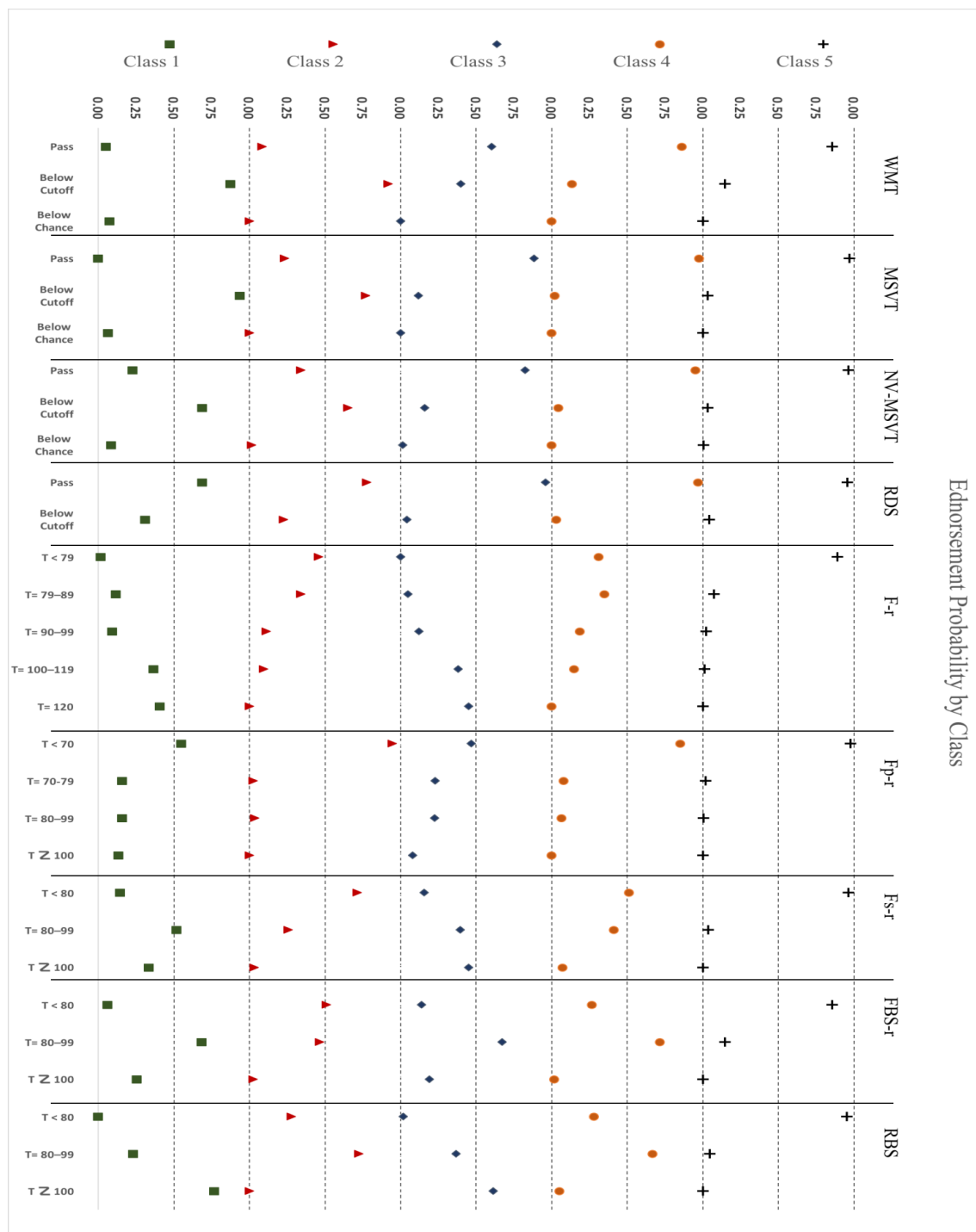
Regarding the SVTs, none were endorsed at the invalid responding level, except for a very small percentage on one SVT (i.e., Fs). Furthermore, on four out of five of the SVT scales, at least 90% of the individuals demonstrated no evidence of over-reporting. For the last SVT scale, FBS-r, 85% of those individuals demonstrated no evidence of over-reporting, a rate significantly higher than any other class.

Table 16.

Conditional Item Probabilities for Five Class Latent Model.

	Class 1	Class 2	Class 3	Class 4	Class 5
WMT					
Pass (0)	0.050	0.083	0.601	0.863	0.855
Below Cut-off (1)	0.874	0.917	0.399	0.137	0.145
Below Chance (2)	0.076	0.000	0.000	0.000	0.000
MSVT					
Pass (0)	0.000	0.232	0.883	0.978	0.969
Below Cut-off (1)	0.936	0.768	0.117	0.022	0.031
Below Chance (2)	0.064	0.000	0.000	0.000	0.000
NV-MSVT					
Pass (0)	0.227	0.338	0.825	0.954	0.965
Below Cut-off (1)	0.687	0.649	0.161	0.046	0.032
Below Chance (2)	0.086	0.013	0.014	0.000	0.003
RDS					
Pass (0)	0.689	0.776	0.960	0.969	0.957
Below Cut-off (1)	0.311	0.224	0.040	0.031	0.043
F-r					
(0) $T < 79$	0.017	0.456	0.000	0.313	0.892
(1) $T = 79 - 89$	0.118	0.338	0.049	0.351	0.072
(2) $T = 90 - 99$	0.091	0.111	0.120	0.187	0.023
(3) $T = 100 - 119$	0.366	0.094	0.382	0.148	0.012
(4) $T = 120$	0.407	0.000	0.449	0.001	0.000
Fp-r					
(0) $T < 70$	0.549	0.943	0.466	0.852	0.976
(1) $T = 70 - 79$	0.160	0.023	0.230	0.079	0.018
(2) $T = 80 - 99$	0.157	0.034	0.224	0.068	0.006
(3) $T \geq 100$	0.135	0.000	0.080	0.000	0.000
Fs					
(0) $T < 80$	0.145	0.713	0.155	0.514	0.964
(1) $T = 80 - 99$	0.519	0.256	0.393	0.412	0.035
(2) $T \geq 100$	0.335	0.031	0.451	0.074	0.001
FBS-r					
(0) $T < 80$	0.062	0.510	0.140	0.267	0.855
(1) $T = 80 - 99$	0.683	0.465	0.671	0.716	0.145
(2) $T \geq 100$	0.255	0.025	0.189	0.017	0.000
RBS					
(0) $T < 80$	0.000	0.278	0.018	0.280	0.954
(1) $T = 80 - 99$	0.232	0.722	0.368	0.667	0.046
(2) $T \geq 100$	0.768	0.000	0.613	0.054	0.000

Figure 1. Conditional Item Probability Profile Plot for Five Class Model.



Note. Y-axis is the probability of endorsing an item; higher endorsement rates.

Table 17.

Number of Criteria Failed and/or Over-endorsed for the Five Class Model.

# of Criteria Failed/Over-endorsed	%	Class 1	Class 2	Class 3	Class 4	Class 5
0	24.6	-	-	-	-	277(53.8)
1	16.1	-	-	-	1 (.4)	180 (35.0)
2	11.1	-	-	1 (.6)	71 (26.9)	53 (10.3)
3	12.6	-	17 (16.7)	6 (3.7)	114 (43.2)	5 (1)
4	11.7	-	27 (26.5)	51 (31.5)	54 (20.5)	-
5	9.1	2 (2.4)	21 (20.6)	56 (34.6)	24 (9.1)	-
6	7.9	18 (21.7)	30 (29.4)	41 (25.3)	-	-
7	3.6	29 (34.9)	5 (4.9)	7 (4.3)	-	-
8	2.4	25 (30.1)	2 (2.0)	-	-	-
9	0.8	9 (10.8)	-	-	-	-
<i>N</i>	1126	83	102	162	264	515
%	100.0	7.4	9.1	14.4	23.4	45.7

Note. Median number of criteria met in each class is bolded.

CHAPTER IV

DISCUSSION

There were two primary goals for this study. The first goal was to examine if validity measures from different domains (i.e., performance, symptom, and embedded performance validity measures) would work in conjunction to identify distinct profiles of malingering. The second goal was to explore the characteristics of these groups and how they differed with respect to performance, symptom, and embedded performance validity measures.

The following discussion is divided into three parts. The first section presents major findings from the first two research questions. The next section examines the implications of these findings with regards to the assessment of performance and symptom validity. The third section is focused on the limitations of this study and suggestions for future research directions for the study of performance and symptom validity.

Latent class analysis was implemented to identify performance and symptom validity profiles that describe different types of malingering. This study used commonly accepted statistical fit criteria to determine best model fit. Unfortunately, fit criteria did not point to one model and suggested a four or five- class model. Subsequently, the conditional item probability plots were examined, and ultimately the five-class model was chosen. The ABIC and LMR pointed to a five-class model and the five-class model made the most sense conceptually. A total of 9 indicators were used in this study. These profiles were then examined and used to categorize different levels of malingering based on the Slick et al. (1999) Malingered Neurocognitive Dysfunction (MND) criteria. The proportion of the performance and symptom validity profiles varied. The *Valid Responders* class had the largest proportion of individuals at 46%, followed by the *Possible Malingering* class at 23%, the *Probable Symptom Malingering* class at 14%, the

Probable Neurocognitive Malingering class at 9%, and the *Definite Malingering* class with the smallest proportion of individuals at 7%.

The results highlight what has been well established regarding malingering research in that measures of performance, symptom, and embedded performance validity are effective in discriminating different types of task engagement and can be used jointly to identify different types of task engagement (Boone, 2008; Larrabee, 2012). The *Definite Malingering* class had the highest level of total failure (both below cut-off and below chance combined) on performance and embedded performance validity measures and had the highest level of invalid responding across most of the SVT scales. Interestingly, there was a near perfect probability of scoring at least below cut-off on the WMT and MSVT. There were still about 23% of individuals who passed the NV-MSVT. Although, these three PVTs were created by the same developer, the NV-MSVT is somewhat different than the WMT and MSVT. The NV-MSVT measures non-verbal memory whereas the WMT and MSVT are verbal memory validity measures (Green, 2003, 2004, & 2008). Additionally, the NV-MSVT has more complex task demands that may make it appear less like a measure of memory (Green, 2008). Therefore, some individuals in the *Definitive Malingering* class who passed the NV-MSVT may have been unlikely to perceive this as a measure to perform poorly. Furthermore, there is greater complexity in the instructions for the NV-MSVT versus the WMT and MSVT (Green, 2010) which may have contributed to test failure.

Compared to the class with the best task engagement, the *Valid Responders* class had the highest passing rates across all validity measures. Again, evidence of invalid responding was minimal in this class for the SVT scales. The stark contrast between these groups is likely due to a deliberate attempt by individuals in the *Definite Malingering* class to perform poorly and

significantly over-report symptoms to portray a high level of impairment. This finding is consistent with research that has previously shown that performance at below chance levels or symptom magnification at the highest levels (i.e., invalid responding) are clear indicators of malingering (Slick et al., 1996). Interestingly, both the *Definitive Malingering* and *Valid Responders* classes were the only two classes that had most likely class membership probabilities above .9. This suggests that for these two classes classification agreement was well above 90% and misclassification rates were low (see Table 15). This is not an unsurprising finding given that identifying response patterns at extremes are often easier to identify. Greater difficulty arises when clinicians attempt to identify response pattern that less obvious or more nuanced.

There were also distinct differences between the *Probable Neurocognitive Malingering* and *Probable Symptom Malingering* classes. The *Probable Neurocognitive Malingering* class had a larger PVT failure rate than the *Probable Symptom Malingering* class, while the latter class inversely had higher rates of SVT over-reporting and invalid responding. This suggests that there was a distinct intention to exaggerate a specific type of symptoms by individuals in these groups. This is also different from the *Definite Malingering* class which is characterized by a global pattern of poor performance and symptom exaggeration. It was also interesting to note that the *Probable Neurocognitive Malingering* class did not produce higher levels on the invalid responding on scales that capture cognitive complaints (i.e., FBS-r and RBS). It appeared that individuals in this class were primarily focused on demonstrating cognitive deficits rather than over-reporting them.

Lastly, the most striking difference between the *Possible Malingering* and *Valid Responders* classes can be seen in the SVT scales. Both classes had near similar passing rates on the PVTs, however, the *Valid Responders* class had only one SVT scale with a 0.1% invalid

responding endorsement probability, whereas four of five SVT scales in the *Possible Malingering* class had 3 scales still demonstrating some invalid responding (albeit below 10%). None of the SVT scales in the *Valid Responders* class reached below 85% valid responding.

An interesting trend among all of the validity measures is that the rate of passing the WMT never reached higher than 86% despite the other two PVT measures approaching near perfect passing rates in the *Valid Responders* and *Possible Malingering* class. This suggests that the WMT was the most sensitive to PVT failure as compared to the MSVT and NV-MSVT. Numerous studies have also shown that the WMT is superior to other PVTs (Mossman, Wygant, & Gervais, 2012). Additionally, RDS failure appeared to be more useful when examining the *Definite Malingering and Probable Neurocognitive Malingering* classes as the RDS failure rate was less than 5% in the *Probably Symptom Malingering, Possible Malingering, and Valid Responders* classes. This is to be somewhat expected give that the RDS is a measure of auditory attentional capacity whereas the three PVTs capture verbal and non-verbal memory. This is significant because the most common complaint in assessments of neurocognitive complaints involve memory as opposed to auditory attention (Vasterling, Brailey, Allain, Duke, Constans, Sutker, 2002; Vasterling & Grailey, 2005).

Implications

This is the first study to use latent class analysis to examine subtypes of malingering while using validity measures from multiple domains. Results from this study indicate that these validity measures work well to distinguish different classes of performance and symptom validity that may be useful in capturing subtypes of malingering. For this study, performance on the validity measures produced 5 distinct subgroups with different patterns of responding. Additionally, the results of this study align well with the conceptual understanding of

malingering (Binder & Willis, 1991; Iverson & Binder, 2000; Larrabee, 1998, 2000; Mittenberg, Rotholz, Russel, & Heilbronner, 1996; Nelson, Sweet & Demakis, 2006; Rogers, 2018; Wygant et al, 2007).

The results of this study also align with the diagnostic criteria for malingered neurocognitive dysfunction (MND; Slick et al., 1999). Similar to the MND criteria, a *Definite Malingering* class emerged that was characterized by a largest percentage of below chance performance. Furthermore, the *Definite Malingering* class largely failed all the PVTs and invalidated the SVT scales which supports research regarding definite noncredible performances (Ashendorf, O'Bryant, & McCaffrey, 2003; Boone 2007; Bush et al., 2005; Iverson & Binder, 2000; Spreen & Strauss, 1998; Thompson, 2002). The *Probable Malingering (Neurocognitive and Symptom)* classes in this study likely represent two subgroups of the Probable Malingering classification proposed in the MND criteria. The *Possible Malingering* class demonstrated some evidence of response bias. The *Valid Responders* class consisted of the largest contingency of individuals who responded validly to these measures. Although there was a very small percentage of failure rate across validity measures, this is not unexpected (Boone, 2009), and suggests these individuals were “honest responders” and demonstrated only an incentive to feign but did not engage in any form of symptom exaggeration or fabrication.

Ultimately, the findings of different subgroups of malingering contribute to the understanding of malingering as a dimensional construct with varying degrees of response patterns rather than a dichotomous perspective (malingering vs. honest responding). This understanding of malingering as a dimensional construct is also supported by Rogers' (2018) description of numerous response styles and motivations for malingering. Several previous studies have attempted to tackle the question of dimensionality regarding malingering and have

shown mixed findings (Frazier et al., 2007; Mossman et al., 2012; Walters et al., 2009; Walters et al., 2009). What cannot be gleaned from the findings of this study, involves the reasons for response bias or reduced task engagement. Some evaluatees may have produced invalid responses due to disengagement from the testing process (i.e., low effort) rather than a deliberate or conscious intention to respond incorrectly. Additionally, some aspects of the MND criteria were not assessed (i.e., symptom report vs. observed behaviors/presentation) with the analyses. Therefore, the results were limited in the ability to identify some malingerers, particularly in the *Possible Malingering* class.

Clinicians and researchers could potentially benefit from this study because this research has identified characteristics associated with different types of malingering, albeit in a single data set. While not exhaustive with the use of only several validity measures, these measures have been well-validated and compared against other PVTs. These results may be helpful for choosing a combination of performance validity tests, as no two PVTs are the same and at times, when one is failed the other is passed (Green, 2007). Research has shown that some PVTs are considerably more sensitive to reduced task engagement than others (Gervais et al., 2004; Green et al., 1999; Green et al., 2001; Mossman et al., 2012), which suggests that evaluatees may escape detection depending on how they respond and which measures they respond to. Furthermore, concerns about valid or invalid responding become increasingly important when questions arise about the accuracy of test scores produced in neuropsychological or forensic evaluations. As stated by the National Academy of Neuropsychology (Bush et al., 2005), “an adequate assessment of response validity is essential in order to maximize confidence in the results of neurocognitive personality measures.”

As stated previously, most studies examining response bias and malingering have used simulation or known groups (criterion) comparison designs. With these research designs, there were disadvantages that limited the utility of their findings. Simulation designs may fail to produce response patterns that are similar to real-world evaluatees (Mossman, Wygant, & Gervais, 2012). Known-group designs are reliant on group assignment criteria which may be imperfect and likely introduces misclassification and misestimation. A prior study by Mossman, Wygant, and Gervais (2012) demonstrated that latent class modeling (LCM) through a Bayesian framework could be implemented to evaluate PVT classification accuracy. This study used data from real-world forensic evaluations to identify underlying groups that tap into the latent structure of malingering. Therefore, this study and the latter examined and based inferences on the behaviors of evaluatees acting under real-world testing conditions and real-world incentives. The LCA method offers a solution to the methodological limitations of known-group and simulation research designs. LCA retains ecological validity (i.e., results from real-world forensic evaluations who have significant motivation, contingencies, and stressors inherent in real evaluations) and removes the need for a perfect truth/exclusion criteria. Despite limitations in these research designs, the results demonstrated here support previous findings described earlier that showed associations between PVTs, SVTs, and structured malingering criteria (Gervais et al., 2008; Gervais et al. 2010; Gervais et al, 2011; Rogers et al., 2011; Schroeder et al., 2012; Tarescavage et al., 2013; Wygant et al., 2009; Wygant et al., 2011; Youngjohn, Wershba, Stevenson, Sturgeon, et al., 2011).

Limitations and Directions for Future Research

There are several limitations with regards to this study. First, the findings from this study are representative of a single data set and a single evaluation context. The above-mentioned

findings are not final judgments about the performance of these measures. Additionally, with regards to LCA, the nominalistic and reification fallacies must be considered. Although, the LCA model demonstrated 5 classes, this is not necessarily evidence that these are real subgroups of malingerers. Additionally, the names of these classes do not make them true. Results may be different if other validity measures are used or if different subjects from different evaluation settings were examined. The scores on these measures used in this study were recoded to capture levels of responding. Therefore, information about particular cut scores was lost with LCA. Latent profile analysis would provide information about the means and variances for respondents in a particular class. More research is needed to replicate the findings from this study.

Covariates were not assessed to examine possible within-group differences between the classes. Lastly, no auxiliary variables were used as predictors or outcomes of the latent classes. Previous research has shown that demographic or psychosocial variables (e.g., age, gender, race/ethnicity, social-economic status), intelligence, and psychopathology do not necessarily impact performance on validity measures; however, exploration of these variables with LCA is warranted. More research is also needed to examine the patterns that may arise within neuropsychological test scores based on class membership.

Another limitation involves the absence of any information external to the PVT, SVT, and EVI data. The statistical approach in this study involved making judgments about response pattern from the data alone. From a clinical context, this is something that most clinical psychologists would not engage in but would rather base their conclusions on a comprehensive evaluation and assessment. This would include obtaining ample data from their clinical encounters with evaluatees, collateral information (e.g., third-party information), behavioral

observations, record reviews, and a synthesis this information with the resultant test data to produce informed clinical determinations.

These measures were administered in a standardized order during the evaluation process. Therefore the order was not randomized have impacted the results due to testing order effects.

Another limitation involves the conditional independence of the observed latent variables in this study given the latent class. High bivariate residuals (pairs with Pearson test statistic > 30) would suggest severe model misfit (Asparouhv & Muthen, 2015). The results showed somewhat high bivariate residuals; however, they were not above the > 30 cutoff. Hageaars and McCutcheon's (2002) have suggested that this may take place when respondents attempt to make their answers consistent with their other responses and that any resultant classes may be a product of this approach. Given this notion, it is not surprising that some conditional dependence may have been observed. The three PVT measures had several differences but followed the same forced-choice recognition paradigm and were created by the same test developer. Therefore, if an individual desired to appear impaired on one forced-choice PVT, that individual more than likely would do the same for other similar PVT measures. As stated previously, future research using LCA modeling may wish to use PVT measures with diverse methods of detections. Measures such as the Rey 15-Item Test (FIT; Rey, 1964), the Dot Counting Test (DCT; Boone, Lu, Back, King et al., 2002; Boone, Lu, & Herzberget, 2002a; Rey, 1941), and the b Test (Boone, Lu, Sherman, Palmer et al., 2000; Boone, Lu, & Herzberg, 2002b) provide an alternative approach to malingering detection through the use of the floor effects principle (Rogers, Harrell, & Liff, 1993). Additionally, this conditional dependence on the SVT scales (e.g., F-r & Fp-r) is understandable, given that the scales come from the same self-report measure and have some similarities between item responses for each scale.

CHAPTER V

CONCLUSIONS

In conclusion, latent class analysis was used to elucidate the latent structure of malingering using a large civil forensic sample of individuals given multiple indicators of response bias. Findings of a five-class model with distinct response patterns demonstrated similarities to what has previously been demonstrated concerning the understanding of malingering. Although, the debate of malingering as a dimensional construct vs. a dichotomy will surely wage on, the findings here ultimately support the use of data points of multiple measures from multiple domains to accurately evaluate response bias.

REFERENCES

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317–332.
- Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *Manual for the Computerized Assessment of Response Bias*. Durham NC: Cognisyst.
- Allport, G. W. (1937). *Personality: A psychosocial interpretation*. New York: Henry Holt.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Pub.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders DSM-IV-TR fourth edition (text revision)*.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed. American Psychiatric Pub.
- Arbisi, P. A., & Ben-Porath, Y. S. (1997). The ability of the Minnesota Multiphasic Personality Inventory-2 validity scales to detect fake-bad responses in psychiatric inpatients. *Psychological Assessment*, 10, 221-228.
- Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, 21, 899–916.
- Archer, R. P., & Smith R. S. (2008). *Personality Assessment*. New York, NY US: Routledge/Taylor & Francis Group.
- Armistead-Jehle, P., & Gervais, R. O. (2011). Sensitivity of the test of memory malingering and the nonverbal medical symptom validity test: A replication study. *Applied neuropsychology*, 18(4), 284-290.

- Armistead-Jehle, P., Gervais, R. O., & Green, P. (2012a). Memory Complaints Inventory results as a function of symptom validity test performance. *Archives of clinical neuropsychology*, 27(1), 101-113.
- Armistead-Jehle, P., Gervais, R. O., & Green, P. (2012b). Memory Complaints Inventory and symptom validity test performance in a clinical sample. *Archives of clinical neuropsychology*, 27(7), 725-734.
- Armistead-Jehle, P., Green, P., Gervais, R. O., & Hungerford, L. D. (2015). An examination of the Word Memory Test as a measure of memory. *Applied Neuropsychology: Adult*, 22(6), 415-426.
- Asparouhov, T., & Muthén, B. (2018). Variable-specific entropy contribution. Retrieved from <http://www.statmodel.com/download/UnivariateEntropy.pdf>
- Asparouhov, T., & Muthén, B. (2015). Residual associations in latent class and latent transition analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(2), 169-177.
- Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 176-189.
- Bagby, R. M., Rogers, R., Buis, T., Nicholson, R. A., Cameron, S. L., Rector, N. A., Schuller, D. R., & Seeman, M. V. (1997). Detecting feigned depression and schizophrenia on the MMPI-2. *Journal of Personality Assessment*, 68(3), 650-664.
- Ben-Porath, Y. S. (2003). Assessing Personality and Psychopathology with Self-Report Inventories. *Handbook of psychology*.
- Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. University of Minnesota Press.

- Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2-Restructured Form): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Ben-Porath, Y. S., & Tellegen, A. (2011). *MMPI-2-RF, Minnesota Multiphasic Personality Inventory-2 Restructured Form: Technical Manual*. University of Minnesota Press.
- Bernard, L. C., McGrath, M. J., & Houston, W. (1996). The differential effects of simulating malingering, closed head injury, and other CNS pathology on the Wisconsin Card Sorting Test: Support for the “pattern of performance” hypothesis. *Archives of Clinical Neuropsychology, 11*, 231-45.
- Berry, D. T. R., Adams, J. J., Clark, C. D., Thacker, S. R., Burger, T. L., Wetter, M. W., Baer, R. A., & Borden, J. W. (1996). Detection of a cry for help on the MMPI-2: An analog investigation. *Journal of Personality Assessment, 67*, 26-36.
- Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W., (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury, 27*, 909-916.
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal, 5*(4), 404-417.
- Bianchini, K. J., Greve, K. W., Love, J. M. (2003). Definite malingered neurocognitive dysfunction in moderate/severe traumatic brain injury. *The Clinical Neuropsychologist, 17*, 574-80.

- Bianchini, K. J., Etherton, J. L., Greve, K. W., Heinly, M. T., & Meyers, J. E. (2008). Classification accuracy of MMPI-2 validity scales in the detection of pain-related malingering: A known groups study. *Assessment, 15*, 435-449.
- Binder, L. M. (1990). Malingering following minor head trauma. *The Clinical Neuropsychologists, 4*, 25-36.
- Binder, L. M., Villanueva, M. R., Howieson, D., & Moore, R. T. (1993). The Rey AVLT recognition memory task measures motivational impairment after mild head trauma. *Archives of Clinical Neuropsychology, 8*, 137-47.
- Binder, L. M. & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*, 175-81.
- Boone, K. B. (Ed.). (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. Guilford Press.
- Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist, 23*, 729-741.
- Boone, K. B. (2013). *Clinical Practice of Forensic Neuropsychology: An Evidence-Based Approach*. Guilford Press.
- Boone, K. B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., Warner-Chacon, K., et al. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The b Test. *Archives of Clinical Neuropsychology, 15*, 227-241.
- Boone, K., Lu, P., & Herzberg, D. S. (2002a). *The Dot Counting Test*. Los Angeles: Western Psychological Services.

- Boone, K., Lu, P., & Herzberg, D. S. (2002b). *The b Test Manual*, Los Angeles: Western Psychological Services.
- Boone, K. B., Lu, P., & Wen, J. (2005). Comparisons of various RAVLT scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology*, *20*, 301-319.
- Breting, L. G., & Sweet, J. (2013). Freestanding cognitive symptom validity tests: Use and selection in mild traumatic brain injury. In *Mild traumatic brain injury, symptom validity assessment and malingering* (pp. 145-158). Springer, New York.
- Bucholz, K. K., Heath, A. C., Reich, T., Hesselbrock, V. M., Krarner, J. R., Nurnberger Jr, J. I., & Schuckit, M. A. (1996). Can we subtype alcoholism? A latent class analysis of data from relatives of alcoholics in a multicenter family study of alcoholism. *Alcoholism: Clinical and Experimental Research*, *20*(8), 1462-1471.
- Bucholz, K. K., Hesselbrock, V. M., Heath, A. C., Kramer, J. R., & Schuckit, M. A. (2000). A latent class analysis of antisocial personality disorder symptom data from a multi-centre family study of alcoholism. *Addiction*, *95*(4), 553-567.
- Bulik, C. M., Sullivan, P. F., & Kendler, K. S. (2000). An empirical study of the classification of eating disorders. *American Journal of Psychiatry*, *157*(6), 886-895.
- Bury, A. S., & Bagby, R. (2002). The detection of feigned uncoached and coached posttraumatic stress disorder with the MMPI-2 in a sample of workplace accident victims. *Psychological Assessment* *14*, 472-484.
- Bush, S. S., Ruff, R. M., Troster, A., Barth, J., Koffler, S. P., Pliskin, N. H., et al. (2005). NAN position paper: Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, *20*, 419-426.

- Butcher, J. N., Arbisi, P. A., Atlis, M. M., & McNulty (2003). The construct validity of the Lees-Haley Fake Bad Scale: Does this measure somatic malingering and feigned emotional distress? *Archives of Clinical Neuropsychology, 18*, 473-485.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). Manual for the administration and scoring of the MMPI-2. *Minnesota: University of Minnesota Press Minneapolis.*
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). Minnesota Multiphasic Personality Inventory-2: Manual for administering, scoring, and interpretation (2nd ed.).
- Chafetz, M. (2011). Reducing the probability of false positives in malingering detection of social security disability claimants. *The Clinical Neuropsychologist, 25*, 1239-1252.
- Chen, L. S., Eaton, W. W., Gallow, J. J., & Nestadt, G. (2000). Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study. *Journal of Affective Disorders, 59*(1), 1-11.
- Conder, R., Allen, L., & Cox, D. (1992). Manual for the computerized assessment of response bias. *Durham, NC, CogniSyst.*
- Croon, M. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology, 43*(2), 171-192.
- Davis, J. J., Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *Clinical Neuropsychologist, 28*, 199-214.
- Dean, A. C., Victor, T., Boone, K. B., Arnold, G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist, 23*, 705-722.

- Dean, A. C., Victor, T., Boone, K. B., Philpott, L., & Hess, R. (2009). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist, 23*, 133-152.
- Delis, D. C., Kramer, J. H. Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult version*. San Antonio, TX: Psychological Corp.
- Delis, D. C., Kramer, J. H. Kaplan, E., & Ober, B. A. (2000). *CVLT-II. California Verbal Learning Test* (2nd ed., adult version). San Antonio, TX: Psychological Corp.
- Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin, 43*, 385–440.
- Etherton, J. L., Bianchini, K. J., Greve, K. W., & Heinly, M. T. (2005). Sensitivity and specificity of reliable digit span in malingered pain-related disability. *Assessment, 12*, 130-36.
- Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using *Mplus*. *Structural Equation Modeling, 18*(1), 132-151.
- Forbes, D., Elhai, J. D., Miller, M. W., & Creamer, M. (2010). Internalizing and externalizing classes in posttraumatic stress disorder: A latent class analysis. *Journal of Traumatic Stress, 23*(3), 340-349.
- Frazier, T. W., Youngstrom, E. Am., Naugle, R. I., Haggerty, K. A., & Busch, R. M. (2007). The latent structure of cognitive symptom exaggeration on the Victoria Symptom Validity Test. *Archives of Clinical Neuropsychology, 22*, 281-300.
- Frederick, R. I. (2000). Mixed group validation: A method to address the limitations of criterion group validation in research on malingering detection. *Behavioral Sciences & the Law, 18*, 693–718.

- Gallagher, R. W. (1997). *Detection of malingering at the time of intake in a correctional setting with the MMPI-2 validity scales*. Unpublished doctoral dissertation, Kent State University, Kent, OH.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment, 14*, 196-208.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2008). Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *The Clinical Neuropsychologist, 22*, 1061–1079.
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Sellbom, M. (2010). Incremental validity of the MMPI-2-RF over-reporting scales and RBS in assessing the veracity of memory complaints. *Archives of Clinical Neuropsychology, 25*, 274–284.
- Gervais, R. O., Green, P., Allen, L. M., & Iverson, G. L. (2001). Effects of coaching on symptom validity testing in chronic pain patients presenting for disability assessments. *Journal of Forensic Neuropsychology, 2*, 1–19.
- Gervais, R. O., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19*(4), 475-487.
- Gervais, R. O., Russell, A. S., Green, P., Allen, L. M., Ferrari, R., & Pieschl, S. D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology, 28*, 1892-1899.
- Giger, P., Merten, T., Merckelbach, H., & Oswald, M. (2010). Detection of feigned crime-related amnesia: A multi-method approach. *Journal of Forensic Psychology Practice, 10*, 440-463.

- Goodwin, B. E., Sellbom, M., & Arbisi, P. A. (2013). Posttraumatic stress disorder in Veterans: The utility of the MMPI-2-RF validity scales in detecting overreported symptoms. *Psychological Assessment, 25*, 671–678.
- Graham, J. R. (2012). *MMPI-2: Assessing personality and psychopathology* (5th ed.). New York, NY: Oxford University Press.
- Granacher, R. P., & Berry, D. T. R. (2008). Feigned medical presentations. In R. Rogers (Ed.) *Clinical assessment of malingering and deception*, 3rd Edition (pp. 145-156). New York: The Guilford Press.
- Green, P. (2003). *Word Memory Test*. Edmonton, Alberta, Canada: Green's Publishing.
- Green, P. (2004). *Medical Symptom Validity Test (MSVT) for Microsoft windows: User's manual*. Paul Green Pub.
- Green, P. (2008). *Manual for the Nonverbal Medical Symptom Validity Test for Windows*. Edmonton, Alberta, Canada: Green's Publishing.
- Green, P., Allen, L. M., & Astner, K. (2003). The word memory test. *Edmonton, Alberta: Green Publishing*.
- Green, P., & Iverson, G. L. (2001). Validation of the computerized assessment of response bias in litigating patients with head injuries. *The Clinical Neuropsychologist, 15*(4), 492-497.
- Green, P., Iverson, G. L., Allen, L. M. (1999) Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13*, 813-819.
- Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*, 1045-1060.

- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Motor dysfunction in profiles in traumatic brain injury and post-concussion syndrome. *Journal of the International Neuropsychological Society, 2*, 477-485.
- Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A. et al. (2006). The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *NeuroToxicology, 27*, 940-50.
- Greve, K. W., Bianchini, K. J., Love, J. M., Brennan, A., & Heinly, M. T. (2006). Sensitivity and specificity of MMPI-2 validity scales and indicators to malingered neurocognitive dysfunction in traumatic brain injury. *Clinical Neuropsychologist, 20*, 491-512.
- Greve, K. W., Bianchini, K. J., Mathias, C. W., Houston, R. J., & Crouch, J. A. (2002). Detecting malingered performance with the Wisconsin Card Sorting Test: A preliminary investigation in traumatic brain injury. *The Clinical Neuropsychologist, 16*, 179-91.
- Greve, K. W., Heinly, M. T., Bianchini, K. J., & Love, J. M. (2009). Malingering detection with the Wisconsin Card Sorting Test in mild traumatic brain injury. *The Clinical Neuropsychologist, 23*, 343-62.
- Greve, K. W., Ord, J. S., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced choice symptom validity tests. *The Clinical Neuropsychologist, 22*, 896-918.
- Greve, K. W., Springer, S., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., et al. (2007). Malingering in toxic exposure. Classification accuracy of Reliable Digit Span and WIAS-III Digit Span scaled scores. *Assessment, 14*, 12-21.
- Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). Hoboken, NJ: Wiley

- Handler, L. & Meyer, G. J. (1998). The importance of teaching and learning personality assessment. In L. Handler & M. J. Hilsenroth (Eds.), *Teaching and learning personality assessment* (pp. 3-30). Mahwah, NJ: Erlbaum.
- Hathaway, S. R., & McKinley, J. C. (1943). Minnesota Multiphasic Personality Inventory; manual (Revised).
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults. *Lutz, FL: Psychological Assessment Resources*.
- Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. J. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*, 892-900.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & the Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.
- Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 16*(3), 257.
- Heinly, M. T., Greve, K. W., Bianchini, K. J. Love, J. M., & Brennan, A. (2005). WAIS Digit Span-based indicators of malingered neurocognitive dysfunction. *Assessment, 12*, 429-44.
- Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 20*, 786-97.

- Hojtink, H. (2001). Confirmatory latent class analysis: Model selection using Bayes factors and (pseudo) likelihood ratio statistics. *Multivariate Behavioral Research*, 36(4), 563-588.
- Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: a meta-analysis of the MMPI-2-Restructured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist*, 1-24.
- Inman, T. H., Vickery, C. D., Berry, D. T., Lamb, D. G., Edwards, C. L., & Smith, G. T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The letter memory test. *Psychological Assessment*, 10(2), 128.
- Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *The Journal of head trauma rehabilitation*, 15(2), 829-858.
- Jones, A., Ingram, M. V., & Ben-Porath, Y. S. (2012). Scores on the MMPI-2-RF scales as a function of increasing levels of failure on cognitive symptom validity tests in a military sample. *The Clinical Neuropsychologist*, 26, 790–815.
- Kane, A. W. Essentials of malingering assessment. *Essentials of forensic psychological assessment* (1999): 78-99.
- Kaufman, D. M., Geyer, H. L., & Milstein, M. J. (2016). Kaufman's Clinical Neurology for Psychiatrists E-Book. Elsevier Health Sciences.
- Keel, P., Fichter, M., Quadflieg, N., Bulik, C., Baxter, M., Thornton, L., et al. (2004). Application of a latent class analysis to empirically define eating disorder phenotypes. *Archives of General Psychiatry*, 61, 192–200.

- Kessler, R. C., Stein, M. B., & Berglund, P. (1998). Social phobia subtypes in the National Comorbidity Survey. *American Journal of Psychiatry*, *155*(5), 613-619.
- Kim, M. S., Boone, K. B., Victor, T., Marion, S. D., Amano, S., Cottingham, M. E., et al. (2010). The Warrington Recognition Memory Test for Words as a measure of response bias: Total score and response time cutoffs developed on “real world” credible and noncredible subjects. *Archives of Clinical Neuropsychology*, *25*(1), 60-70.
- Kim, N., Boone, K. B., Victor, T., Lu, P., Keatinge, C., & Mitchell, C. (2010). Sensitivity and specificity of a Digit Symbol recognition trial in the identification of response bias. *Archives of Clinical Neuropsychology*, *25*, 420-428.
- Lanyon, R. I. (1970). Development and validation of a psychological screening inventory. *Journal of Consulting and Clinical Psychology*, *35*(Suppl. 1, Pt. 2), 1–24.
- Lanyon, R. I. (1993). Development of scales to assess specific deception strategies on the Psychological Screening Inventory. *Psychological Assessment*, *5*, 324–329.
- Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in litigating subjects. *The Clinical Neuropsychologist*, *12*, 179-88.
- Larrabee, G. J. (2000). Forensic neuropsychological assessment. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (2nd ed., pp. 301-35). Mahwah, NJ: Lawrence Erlbaum.
- Larrabee, G. J. (2003a). Detection of malingering using atypical performance patterns on standard neuropsychological test, *The Clinical Neuropsychologist*, *17*, 410-425.
- Larrabee, G. J. (2003b). Exaggerated MMPI-2 symptom report in personal injury litigants with malingered neurocognitive deficit. *Archives of Clinical Neuropsychology*, *18*, 673-86.

- Larrabee, G. J. (2008). Aggregation across indicators improves the detection of malingering: relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*, 666-679.
- Larrabee, G. J. (2009). Malingering scales for the Continuous Recognition Memory Test and Continuous Visual Memory Test, *The Clinical Neuropsychologists, 23*, 167-180.
- Larrabee, G. J. (2012). Assessment of malingering. In Larrabee (Eds.), *Forensic neuropsychology: A scientific approach* (pp. 116-159). Oxford University Press.
- Larrabee, G. J. (2014). Minimizing false positive error with multiple performance validity tests: Response to Bilder, Sugar, and Helleman. *The Clinical Neuropsychologist, 28*, 1230-1242.
- Larrabee, G. J. (2015a). Performance and symptom validity: A perspective from the adult literature. In Kirkwood (Eds.), *Validity testing in child and adolescent assessment: Evaluating exaggeration, feigning, and noncredible effort* (pp. 62-76). Guilford Press.
- Larrabee, G. J. (2015b). The multiple validities of neuropsychological assessment. *American Psychologist, 70*(8), 779-778.
- Laudy, O., Boom, J., & Hoijtink, H. (2005). Bayesian computational methods for inequality constrained latent class analysis. *New developments in categorical data analysis for the social and behavioral sciences*, 63-82.
- Lee, T. T., Graham, J. R., Sellbom, M., & Gervais, R. O. (2012). Examining the potential for gender bias in the prediction of symptom validity test failure by MMPI-2 symptom validity scale scores. *Psychological assessment, 24*(3), 618.
- Lees-Haley, P. R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology, 53*(7), 745-755.

- Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal-injury claimants. *Psychological Reports, 68*, 203-210.
- Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1996). Forensic neuropsychological test usage: An empirical survey. *Archives of Clinical Neuropsychology, 11*(1), 45-51.
- Marion, B. E., Sellbom, M., & Bagby, R. M. (2011). The detection of feigned psychiatric disorders using the MMPI-2-RF overreporting validity scales: An analog investigation. *Psychological Injury and Law, 4*, 1–12.
- McBride, W. F., Crighton, A. H., Wygant, D. B., & Granacher, R. P. (2013). It's Not All in Your Head (or at Least Your Brain): Association of traumatic brain lesion presence and location with performance on measures of response bias in forensic evaluation. *Behavioral sciences & the law, 31*(6), 779-788.
- McCutcheon, A. L. (1987). *Latent class analysis* (No. 64). Sage.
- McLachlan, G., & Peel, D. (2000). Mixtures of factor analyzers. In *In Proceedings of the Seventeenth International Conference on Machine Learning*.
- Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*(4), 266.
- Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*.
- Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports, 78*(3 suppl), 1091-1227.

- Merten, T., Green, P., Henry, M., Blaskewitz, N., & Brockhaus, R. (2005). Analog validation of German-language symptom validity tests and the influence of coaching. *Archives of Clinical Neuropsychology*, *20*(6), 719-726.
- Meyers, J. E., Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, *18*, 261-276.
- Miller, H. A., (2001). *Miller-Forensic Assessment of Symptoms Test (M-FAST): Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Miller, L. S., & Rohling, M. L. (2001). A statistical interpretive method for neuropsychological test data. *Neuropsychology Review*, *11*, 143–169.
- Millis, S. R. Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment*, *7*, 463-471.
- Millon, T. (1997). *Millon Clinical Multiaxial Inventory-III [Manual Second Edition]*. Bloomington, MN: Pearson Assessments.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094-1102.
- Mittenberg, W., Rotholz, A., Russel, E., & Heilbronner, R. (1996). Identification of malingered head injury on the Halstead-Reitan Battery. *Archives of Clinical Neuropsychology*, *11*, 271-81.
- Morey, L. C. (1991). *The Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

- Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources.
- Mossman, D., Wygant, D. B., & Gervais, R. O. (2012). Estimating the Accuracy of Neurocognitive Effort Measures in the Absence of a “Gold Standard. *Psychological Assessment*, 24(4), 815 – 822.
- Mossman, D., Miller, W. G., Lee, E. R., Gervais, R. O., Hart, K. J., & Wygant, D. B. (2015). A Bayesian approach to mixed group validation of performance validity tests. *Psychological assessment*, 27(3), 763.
- Muthén, B. (2004). Latent variable analysis. *The Sage handbook of quantitative methodology for the social sciences*. Thousand Oaks, CA: Sage Publications, 345-68.
- Muthen, L. K., & Muthen, B. O. (2009). Categorical Latent Variable Modeling Using Mplus: Cross-Sectional Data [Powerpoint slides & Lecture] Retrieved from Mplus Short Courses: http://www.statmodel.com/course_materials.shtml
- Muthén, L. K., & Muthén, B. (2015). *Mplus user's guide* (5th ed.). Los Angeles: Muthen & Muthen.
- Nathan, P. E., & Langenbucher, J. (2003). Diagnosis and classification. *Handbook of Psychology*, 1-26.
- Nelson, N. W., Sweet, J. J., & Demakis, G. J. (2006). Meta-analysis of the MMPI-2 Fake Bad Scale: Utility in forensic practice. *The Clinical Neuropsychologist*, 20,39-58.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, 14(4), 535-569.

- Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills, 57*, 367-372.
- Pankratz, L., Fasti, S., & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology, 43*, 421-422.
- Paulhus, D. L. (1988). Balanced inventory of desirable responding (BIDR). *Acceptance and Commitment Therapy. Measures Package, 41*.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.
- PLCH, A. P. L. C. H., & McCormack, F. X. (2010). ICD-10-CM.
- Rasmussen, E., Neuman, R., Heath, A., Levy, F., Hay, D., & Todd, R. (2002). Replication of the latent class structure of attention-deficit/hyperactivity disorder (ADHD) subtypes in a sample of Australian twins. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*, 1018–1028.
- Rawling, P., & Brooks, N. (1990). Simulation index: A method for detecting factitious errors on the WAIS-R and WMS. *Neuropsychology, 4*, 234-238.
- Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment, 10*(1), 10.
- Richman, J., Green, P., Gervais, R., Flaro, L., Merten, T., Brockhaus, R., & Ranks, D. (2006). Objective tests of symptom exaggeration in independent medical examinations. *Journal of Occupational and Environmental Medicine, 48*(3), 303-311.

- Rogers, R. (1990a). Models of feigned mental illness. *Professional Psychology: Research and Practice*, 21, 182-188.
- Rogers, R. (1990b). Development of a new classificatory model of malingering. *Bulletin of the American Academy of Psychiatry and Law*, 18, 323-333.
- Rogers, R. (4th Ed.). (2018). *Clinical assessment of malingering and deception*. Guilford Press.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structure Interview of Reported Systems*. Odessa, FL: Psychological Assessment Resources.
- Rogers, R., & Bender, S. (2003). Evaluation of malingering and deception. In A. M. Goldstein & I. B. Weiner (Eds.), *Handbook of psychology: Vol. 11. Forensic psychology* (pp. 109–129). New York, NY: Wiley.
- Rogers, R., Gillard, N. D., Berry, D. T. R., & Granacher, R. P. (2011). Effectiveness of the MMPI-2-RF Validity Scales for feigned mental disorders and cognitive impairment: A known-groups study. *Journal of Psychopathology and Behavioral Assessment*, 33, 355–367.
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *SIRS-2: Structured Interview of Reported Symptoms: Professional manual*. Psychological Assessment Resources, Incorporated.
- Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders a meta-analysis of the MMPI-2 and malingering. *Assessment*, 10(2), 160-177.
- Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, 67,629-640.
- Rogers, R., Sewell, K. W., Ustad, K. L. (1995). Feigning among chronic outpatients on the MMPI-2: A systematic examination of fake-bad indicators. *Assessment*, 2, 81-89.

- Ross, S. R., Millis, S. R., Krukowski, R. A., Putnam, S. H., & Adams, K. M. (2004). Detecting incomplete effort on the MMPI-2: An examination of the fake-bad scale in mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *26*, 115–124.
- Rubinsky, E. W., & Brandt, J. (1986). Amnesia and criminal law: A clinical overview. *Behavioral Sciences and the Law*, *4*, 27-46.
- Ruscio, J., Haslam, N. & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide*. Mahwah, N. J.: Lawrence Erlbaum Associates.
- Schroeder, R. W., & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations, *The Clinical Neuropsychologist*, *25*, 437-453.
- Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.
- Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343.
- Sellbom, M., & Bagby, R. M. (2010). Detection of overreported psychopathology with the MMPI-2 RF Form Validity Scales. *Psychological Assessment*, *22*, 757–767.
- Sellbom, M., Lee, T. T., Ben-Porath, Y. S., Arbisi, P. A., & Gervais, R. O. (2012). Differentiating PTSD symptomatology with the MMPI-2-RF (Restructured Form) in a forensic disability sample. *Psychiatry Research*, *197*(1-2), 172-179.
- Sellbom, M., Wygant, D. B., & Bagby, M. (2012). Utility of the MMPI-2-RF in detecting non-credible somatic complaints. *Psychiatry Research*, *197*, 295–301.
- Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to

- neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology*, 18(6), 911-922.
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545-561.
- Smart, C. M., Nelson, N. W., Sweet, J. J., Bryant, F. B., Berry, D. T. R., Granacher, R. P., & Heilbronner, R. L. (2008). Use of MMPI-2 to predict cognitive effort: A hierarchically optimal classification tree analysis. *Journal of the International Neuropsychological Society*, 14, 842-852.
- Sollman, M. J., Ranseen, J. D., Berry, D. T. R. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, 22, 325-335.
- Strong, D. R., Glassmire, D. M., Frederick, R. I., & Greene, R. L. (2006). Evaluating the latent structure of the MMPI-2 F (p) scale in a forensic sample: a taxometric analysis. *Psychological assessment*, 18(3), 250.
- Strong, D. R., Greene, R. L., Schinka, J. A. (2000). A taxometric analysis of MMPI-2 infrequency scales [F and F(p)] in clinical settings. *Psychological Assessment*, 12, 166-73.
- Suhr, J. A. & Boyer, D. (1999). Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. *Journal of Clinical and Experimental Neuropsychology*, 21, 701-708.
- Suhr, J., Tranel, D., Wefel, J., & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology*, 19, 500-514.

- Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist, 16*(4), 495-505.
- Tarescavage, A. M., Wygant, D. B., Gervais, R. O., & Ben-Porath, Y. S. (2013). Association between the MMPI-2 Restructured Form (MMPI-2-RF) and malingered neurocognitive dysfunction among nonhead injury disability claimants. *The Clinical Neuropsychologist, 27*, 313–335.
- Tenhula, W. N. & Sweet, J. J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *The Clinical Neuropsychologist, 10*, 104-116.
- Thomas, M. L., Lanyon, R. I., & Millsap, R. E. (2009). Validation of diagnostic measures based on latent class analysis: A step forward in response bias research. *Psychological Assessment, 21*(2), 227.
- Tombaugh, T. N. (1996). Test of memory malingering: TOMM. New York/Toronto: MHS.
- Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*(3), 260.
- van der Werf, S. P., Prins, J. B., Jongen, P. J., van der Meer, J. W., & Bleijenberg, G. (2000). Abnormal neuropsychological findings are not necessarily a sign of cerebral impairment: a matched comparison between chronic fatigue syndrome and multiple sclerosis. *Cognitive and Behavioral Neurology, 13*(3), 199-203.
- Vasterling, J. J., & Brailey, K. (2005). Neuropsychological findings in adults with PTSD. In J. J. Vasterling & C. R. Brewin (Eds.), *Neuropsychology of PTSD: Biological, cognitive, and clinical perspectives* (pp. 178 – 207). New York: Guilford Press.

- Vasterling, J. J., Brailey, K., Allain, A. N., Duke, L. M., Constans, J. I., & Sutker, P. B. (2002). Attention, learning, and memory performances and intellectual resources in Vietnam veterans: PTSD and no disorder comparisons. *Neuropsychology, 16*, 5 – 14.
- Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the Meaning of Multiple Symptom Validity Test Failure. *The Clinical Neuropsychologist, 23*, 297-313.
- Wall, T. D., Wygant, D. B., & Gallagher, R. W. (2015). Identifying overreporting in a correctional setting: Utility of the MMPI-2 restructured form validity scales. *Criminal Justice and Behavior, 42*, 610–622.
- Walters, G. D., Berry, D. T. R., Lanyon, R. I., & Murphy, M. P. (2009). Are exaggerated health complaints continuous or categorical? A taxometric analysis the Health Problem Overstatement Scale. *Psychological Assessment, 21*, 219 - 226.
- Walters, G. D., Berry, D. T. R., Rogers, R., Payne, J. W., & Granacher, R. P. (2009). Feigned neurocognitive deficit: Taxon or dimension? *Journal of Clinical and Experimental Neuropsychology, 31*, 584-93.
- Walters, G. D., Rogers, R., Berry, D. T. R., Miller, H. A., Duncan, S. A., McCusker, P. J., et al. (2008) Malingering as a categorical or dimensional construct: The latent structure of feigned psychopathology as measured by the SIRS and MMPI-2. *Psychological Assessment, 20*, 238-47.
- Wechsler, D. (2008). *WAIS-IV. Administration and scoring manual*. San Antonio, TX: Pearson.
- Whitney, K. A., Davis, J. J., Shepard, P. H., & Herman, S. M. (2008). Utility of the Response Bias Scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Archives of Clinical Neuropsychology, 23*, 777–786.

- Widows, M. R., & Smith, G. P. (2005). SIMS-Structured Inventory of Malingered Symptomatology. *Professional manual*. Lutz, FL: PAR: Psychological Assessment Resources.
- Wiggins, E. C., & Brandt, J. (1988). The detection of simulated amnesia. *Law and Human Behavior, 12*, 57-78.
- Wiggins, C. W., Wygant, D. B., Hoelzle, J. B., & Gervais, R. O. (2012). The more you say the less it means: Overreporting and attenuated criterion validity in a forensic disability sample. *Psychological Injury and Law, 5*(3-4), 162-173.
- Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J., & Sweet, J. J. (2010). Effort indicators within the California Verbal Learning Test-II (CVLT-II). *The Clinical Neuropsychologist, 24*, 153-68.
- Wygant, D. B., Anderson, J. L., Sellbom, M., Rapier, J. L., Allgeier, L. M., & Granacher, R. P. (2011). Association of the MMPI-2 Restructured Form (MMPI-2-RF) Validity scales with structured malingering criteria. *Psychological Injury and Law, 4*, 13–23.
- Wygant, D. B., Ben-Porath, Y. S., & Arbisi, P. A. (2004). Development and initial validation of a scale to detect somatic over-reporting. In *39th Annual Symposium on Recent Developments in the Use of the MMPI-2 and MMPI-A, Minneapolis, MN*.
- Wygant, D. B., Ben-Porath, Y. S., Arbisi, P. A., Berry, D. T. R., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 restructured form (MMPI-2-RF) validity scales in civil forensic settings: Findings from simulation and known group samples. *Archives of Clinical Neuropsychology, 24*, 671–680.
- Wygant, D. B., Sellbom, M., Gervais, R. O., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2010). Further validation of the MMPI-2 and MMPI-2-RF

Response Bias Scale: Findings from disability and criminal forensic settings. *Psychological Assessment*, 22(4), 745.

Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. I. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology*, 22, 488-499.

Youngjohn, J. R., Wershba, R., Stevenson, M., Sturgeon, J., & Thomas, M. L. (2011). Independent validation of the MMPI-2-RF somatic/cognitive and Validity Scales in TBI litigants tested for effort. *The Clinical Neuropsychologist*, 25, 463–476.

VITA

Willie Floyd McBride III

726 Cecil Ave, Louisville, KY, 40211

EDUCATION

- 2007-11 Bachelor of Art, University of Louisville, Psychology
 2011-13 Master of Science, Eastern Kentucky University, Clinical Psychology
 2013-18 Doctor of Philosophy, Virginia Consortium Program in Clinical Psychology

CLINICAL EXPERIENCE

- 2017-18 Vanderbilt – Department of Veteran Affairs Internship in Professional Psychology
 2018-19 Tennessee Valley Healthcare System Fellowship in Interprofessional Psychology

PROFESSIONAL EXPERIENCE

- 2012-13 Research Assistant, Eastern Kentucky University
 2013-14 Adjunct Faculty, Abnormal Psychology, Norfolk State University
 2014-15 Graduate Assistant for Office of Intercultural Relations at Old Dominion University,
 Black Student initiatives
 2018-19 Adjunct Faculty, Forensic Psychology, Eastern Kentucky University

PUBLICATIONS & PRESENTATIONS

- McBride III, W. F.**, Crighton, A. H., Wygant, D. B., & Granacher, R. P. (2013). It's not all in Your Head (or at Least Your Brain): Association of Traumatic Brain Lesion Presence and Location with Performance on Measures of Response Bias in Forensic Evaluation. *Behavioral Science & the Law*, 31, 779-788. doi: [10.1002/bsl.2083](https://doi.org/10.1002/bsl.2083)
- McBride, W. F.**, LaTulip, S. E., Handel, R. W., Brokenborough, M. A., Arbisi, P. A., Wygant, D. B. (2016, May). Examining the Measurement Invariance of the Minnesota Multiple Personality Inventory-2 Restructured Form Externalizing Specific Problems Scales in African American and Caucasian Men. Paper presentation: 2016 MMPI 51st Annual Symposium, Hollywood, FL.
- McBride, W. F.**, Wygant, D.B. (2016, July). The MMPI-2-RF Validity Scales: Review and Use in Forensic Psychological Assessment. Paper presentation: 2016 European Association of Psychology & Law, Toulouse, France.
- McBride, W.F.**, Sutherland, H.G., Wygant, D.B., & Granacher, R.P. (2013, March). Comparability of SIRS and SIRS-2 Malingering Classifications with External Response Bias Criteria. Poster presentation: 2013 American Psychology Law Society Annual Conference, Portland, OR.