

Eastern Michigan University

DigitalCommons@EMU

Master's Theses and Doctoral Dissertations

Master's Theses, and Doctoral Dissertations,
and Graduate Capstone Projects

2022

A comparison of forensic populations using the validity scales on the MMPI-3

Steven W. Steinert

Follow this and additional works at: <https://commons.emich.edu/theses>



Part of the [Clinical Psychology Commons](#)

A Comparison of Forensic Populations Using the Validity Scales on the MMPI-3

by

Steven W. Steinert

Dissertation

Submitted to the College of Arts and Sciences

Eastern Michigan University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Clinical Psychology

Dissertation Committee:

Eamonn P. Arble, PhD

Tamara Loverich, PhD

Chong Man Chow, PhD

Nicole Kletzka, PhD

Yossef S. Ben-Porath, PhD, ABPP

May 10, 2022

Ypsilanti, Michigan

Acknowledgments

There are so many individuals who have contributed to my career to whom I would like to express my sincere gratitude. First, I would like to thank several individuals from UW-Oshkosh who provided foundational training, inspiration, and opportunities that shaped my early career aspirations and helped launch my career. I thank Phan Y. Hong, PhD, David A. Lishner, PhD, and Quin M. Chrobak, PhD for their supervision and guidance throughout my bachelor's and master's degrees at UW-Oshkosh. I am grateful for their dedication to my success during the early years of my education. I also thank Michael J. Vitacco, PhD, ABPP from Augusta University for his consultation, encouragement, and continuous guidance throughout the different stages of my training. I also am grateful to Sneha Shankar, MS for her statistical consultation, collaboration, and friendship throughout graduate school. Her presence as a lab-mate made my experience in graduate school more fruitful and enjoyable and I could not have been successful without her. Further, I thank Yossef S. Ben-Porath, PhD, ABPP of Kent State University for the opportunity to contribute to and participate in his research on the MMPI-3 and for his participation and consultation on this project. I am honored to have had the opportunity to work on such important research and to learn from his profound expertise on this project as well as others. I am also grateful to Chong Man Chow, PhD and Tamara Loverich, PhD from Eastern Michigan University for lending their expertise and their continued dedication as committee members on this dissertation. Further, I thank Nicole Kletzka, PhD for the many valuable opportunities and experiences she provided me at the Center for Forensic Psychiatry, her ongoing support and encouragement, and for contributing her expertise to this dissertation. I also thank Eamonn P. Arble, PhD for his mentorship throughout this dissertation and throughout my doctoral education. I attribute much of my success personally and professionally to his expertise

and dedication as a psychologist, teacher, mentor, and friend. He sets an example of success and quality that one can only hope to emulate. I am honored to have had the opportunity to work for and along side him.

I would be remiss if I did not thank the many family and friends who provided support and inspiration throughout this process. Although I name several, there are many more friends and family without whom I would not be where I am today. First, I remember and thank my grandfather, Dr. David L. Bowman, for inspiring me toward achieving academic excellence through his own career in higher education. He set an example early in my life of the profound impact one can have on society and other people when they pursue and embrace higher education. I also thank my brother, Nate Steinert, and my sister, Kimmy Steinert, for their support and friendship throughout my life. I am proud of the good people they both have become and I strive to keep up with the high bar for integrity and excellence they both set each day. I also thank my parents, Bill and Carol Steinert, for their support throughout my life and education. Through their own lives and careers, they have instilled in me and have exemplified strong moral values, a dedicated work ethic, compassion for others, and a devotion to improving the lives of others. I am grateful for all they have sacrificed and provided for me, and I hope that I can use all of my talents, knowledge, and gifts to follow their example by creating a better world for others. Finally, I thank my wife, Molly Steinert. My life would truly be incomplete and I would not be half the person I am today without her by my side. Without her, life would have little to no meaning. She has stood by me throughout the entirety of my education and training, maintained a positive, open, and flexible approach to new challenges, and served as a constant source of love and support throughout the challenges of graduate school. For all of that, I am eternally grateful.

I am humbled to consider all of the important people who have contributed so much and remained dedicated to my happiness and success. I will carry the lessons that each person has provided me throughout my life and career. It will be my goal to emulate their personal and professional examples in the work that I do each day, and the values and knowledge they have instilled in me will help me to continue to pursue our united quest to create a better world for us all.

Abstract

Examining the impact of different factors influencing the validity of an individual's self-report during a psychological assessment is important in ensuring valid clinical findings and useful recommendations. These factors are often referred to as response biases. There are multiple types of response bias that can negatively influence the validity of self-reports in clinical assessment contexts. Specifically, individuals undergoing an assessment can be impacted by non-content-based response bias, overreporting, and underreporting of psychological impairment and/or distress. Minnesota Multiphasic Personality Inventory (MMPI) instruments are amongst the leading tools within professional psychology used to identify response bias. The most recent iteration, the MMPI-3, incorporates the latest updates and normative comparison data into validity scales that are designed to capture the different domains of response bias. The current study sought to explore and identify different clinical and contextual factors that influence response bias amongst different groups of individuals involved in a forensic system using the MMPI-3. The results suggest that both situational context as well as psychological impairment and distress may play roles in different levels and types of response bias between different groups of people within a forensic system. This study serves as an important first step in better identifying the unique threats to assessment validity amongst different individuals involved in forensic systems.

Table of Contents

Acknowledgments	ii
Abstract.....	v
List of Tables	ix
Introduction	1
Response Bias.....	1
History of Self-Report Measures	3
MMPI Validity Scales	9
MMPI-2 Validity Scales.....	13
MMPI-2-RF Revision and Validity Scales.....	15
Cannot Say Scale	16
Variable Response Inconsistency-revised (VRIN-r)	17
True Response Inconsistency-revised (TRIN-r).....	17
Infrequency-revised (F-r)	17
Infrequency-psychopathology (Fp-r).....	18
Infrequent Somatic Responses (Fs-r)	18
Symptom Validity Scale (FBS-r)	18
Response Bias Scale (RBS).....	19
Uncommon Virtues (L-r).....	19
Adjustment Validity Scale (K-r).....	19
MMPI-2-RF-EX & MMPI-3	19
Non-Content-Based Validity Scales	20
Overreporting Validity Scales	20

Underreporting Validity Scales	21
Forensic Applications and Legal Standing of the MMPI	22
Forensic Populations.....	25
Incompetent to Stand Trial (IST).....	25
Not Guilty by Reason of Insanity (NGRI).....	28
Use of MMPI Validity Scales With Inpatient and Forensic Populations ..	32
CST & Non-Content-Based Validity Scales	32
CNS Scale.....	33
VRIN & TRIN	33
Influences on Inconsistent Responding	36
Overreporting (F, Fb, Fp, Fs, FBS, RBS).....	37
F and Fb Scales.....	37
Fp Scale	39
Fs Scale.....	41
FBS Scale	43
RBS Scale.....	45
Overreporting and Evaluee/Inpatient Status.....	47
Underreporting Scales (L & K)	48
Underreporting and Evaluee/Inpatient Status.....	49
Gaps in the Literature	51
Hypotheses.....	55
Terminology	55
Hypothesis 1	56

Hypothesis 2	57
Hypothesis 3	57
Hypothesis 4	57
Method.....	59
Data, Data Storage, and Variables.....	59
Population and Setting.....	59
Informed Consent	62
Procedure	62
Statistical Analyses.....	64
Hypothesis 1	65
Hypothesis 2	65
Hypothesis 3	66
Hypothesis 4	66
Results	67
Hypothesis 1	68
Hypothesis 2	69
Hypotheses 3 & 4	72
Discussion.....	74
Conclusions	87
Limitations.....	89
Future Directions	91
References	94

List of Tables

Table 1 – Demographic Frequencies	67
Table 2 – Descriptive Means	68
Table 3 – Group Means and Standard Deviations	68
Table 4 – Independent Samples t-tests between Inpatients and Evaluatees	73

Introduction

Response Bias

Forensic evaluations are conducted to provide information regarding the state of mind of an individual charged with a crime prior to or following treatment and a judicial finding.

Forensic assessments are conducted to inform about the state of mind of an individual undergoing treatment following a judicial finding. Though individuals within the legal system who undergo psychological evaluations likely share some characteristics with others engaged in various forensic contexts, forensic populations are diverse. Forensic evaluatee and inpatient populations each consist of individuals with varying types and degrees of psychopathology as well as external pressures. Researchers have examined unique qualities of these individuals to better understand distinguishing factors related to subgroups of forensic populations.

The nature of the adversarial judicial system inherently presupposes some questioning of an evaluatee's credibility (Rogers & Payne, 2006). Many individuals who undergo psycholegal evaluations experience external pressures that contribute to a unique presentation. This external pressure may cause individuals to inaccurately represent themselves such as to influence the outcome of a psycholegal evaluation. Further, Rogers and Bender (2003) found that internal pressures, such as psychological disorders, identity, or intentional goals, might also impact the accuracy of reports given by individuals undergoing psychological evaluation within the legal system. These responses to internal and external pressures within a psychological evaluation are conceptualized as *response biases*. Response biases can impact a person's ability or desire to provide accurate information pertinent to a forensic evaluation.

There are several types of response biases that are important to consider within a forensic psychological evaluation. In the latest edition of a seminal text on psychological measurement of

response style and malingering, Rogers (2018) provided a comprehensive examination of different response styles that are commonly found in an evaluative setting. Specifically, he identifies three primary response styles that are important for the present research.

First, people can intentionally or unintentionally overreport the presence of psychopathology (Rogers, 2018). Overstated psychopathology can occur due to an overt effort to feign psychological impairment or dysfunction. This response style is sometimes referred to as faking bad (Rogers, 2018).

Additionally, Rogers (2018) indicated that in contrast to overstated psychopathology, people can engage in simulated adjustment wherein they respond such as to “fake good” by either denying symptoms (e.g., “I never feel sad”) or by endorsing unlikely virtues (e.g., “I have never used a swear word”). Individuals being evaluated can be influenced to provide an overly virtuous or symptom-free presentation that is not reflective of their true functioning in different forensic contexts (Melton et al., 2018). Rogers (2018) indicated that although individuals involved in civil litigation are particularly likely (and have received the most research attention) to engage in faking good, individuals undergoing any high-stakes evaluation may be incentivized to present themselves in an overly positive light.

Rogers (2018) also reported that people can engage in response styles irrespective of the content of evaluative questions. Specifically, they can engage in irrelevant responding, random responding, acquiescent responding, or disacquiescent responding. This response style is sometimes referred to as random responding and yea-saying or nay-saying.

Individuals in forensic contexts are likely to experience multiple and varied internal and external pressures that can impact response style (Rogers & Payne, 2006). An individual’s response style is, in part, likely impacted by their stage in the legal process and their desired

outcome related to their legal situation. It is likely that those who receive certain forensic opinions have specific clinical presentations that could also contribute to certain types of response biases. Evaluatees who undergo a forensic evaluation for criminal responsibility (CR) and/or competency to stand trial (CST) and forensic inpatients are likely to have distinguishing response biases due to the nature of their situational factors (Graham, 2006).

The different pressures and related response biases have been researched and identified using various psychological assessment instruments. Given that forensic assessments frequently rely on self-reports provided by test-takers, researchers have developed self-report measurement tools to evaluate the aforementioned response biases (Wygant & Granacher, 2015).

Pertinent to the current research, Minnesota Multiphasic Personality Inventory (MMPI) instruments have been used to provide data points regarding response biases within forensic evaluations. The different MMPI iterations have been among the most popular instruments used to examine response bias, and they are commonly used to aid in forensic evaluations examining CST and CR. The latest edition of the MMPI, the MMPI-3, which was recently released, includes the most up-to-date scales examining the aforementioned response bias domains. Given the recent release of the MMPI-3 and the importance of MMPI validity scales in forensic evaluations, it is important to study response styles that are common among different forensic populations. The forthcoming research will provide a framework that will guide hypotheses about the impact of psychiatric and situational influences on response bias amongst forensic evaluatees and inpatients.

History of Self-Report Personality Measures

The development of psychology as a scientific discipline arose out of a curiosity about the specific underlying mechanisms and traits that make people unique. Researchers have long

sought answers to questions about why individuals react or behave differently in a variety of situations. As the development of standardized procedures to assess psychological functioning arose in the early twentieth century, methods to provide valid explanatory profiles improved. Despite the interest of scholars who are often dictated by intrinsic curiosity, understanding human psychological functioning has been important in practical contexts as well.

The onset of World War I prompted a need for efficient yet effective psychological assessment tools that could be used to screen large numbers of individuals (Greene, 2011). Military psychologists touted the potential utility of such measures in selection of high-ranking officials or to determine whether individuals were emotionally fit for various types of battle (Super, 1949). As such, an early edition of the Woodsworth Personal Data Sheet (Woodsworth, 1917) was a personality assessment designed as a scientifically rooted measure of personality, partially in response to the development of standardized assessments for measuring intelligence and calls for psychological assessments to have quantitative foundations (Gabby & Zicker, 2008). As World War I intensified, the need for assessment measures to examine different domains of psychological functioning increased. The development of advanced military technology led to weapons that caused psychological injuries (Gabby & Zicker, 2008). It was estimated that well over one million soldiers suffered from symptoms of nausea, heart palpitations, weeping, and amnesia following bombardment (Hale, 1995), leading to the development of the term *shell-shocked* to describe their psychological condition.

Given the large number of soldiers impacted by shell-shock, the United States military funded the development of an assessment that could be used to determine an individual's susceptibility to the condition. Subsequently, the Scale of Psychoneurotic Tendencies (Woodsworth, 1919) was developed as a measure of emotional stability. The measure was

implemented to screen for individuals who were potentially psychologically vulnerable to shellshock. Following the screening, individuals flagged by the measure were evaluated more comprehensively by military psychiatrists to determine whether they would be permitted to serve, and in what capacity (Greene, 2011). Gabby and Zicker (2008) highlighted that following World War I, Woodsworth renamed his measure, the Woodsworth Personal Data Sheet and began to apply it in private industry settings.

Following the developments in personality assessment stimulated by World War I, and the potential for future utility, researchers continued to develop innovative techniques for measuring different domains of personality in various contexts (Greene, 2011). Early measures of personality, such as the Woodsworth Personal Data Sheet, were comprised of only one dimension of psychological functioning. Improvements in methodology and theoretical advancement in assessment development contributed to the development of multidimensional assessment measures. The Bernreuter Personality Inventory (BPI) consisted of different domains of personality, which initially included neuroticism, dominance, introversion, and self-sufficiency. Conscientiousness and solitariness were later added to a more refined BPI measure (Bernreuter, 1935). Given its multidimensional structure, the BPI became a popular assessment measure in a variety of settings (Super, 1942). Specifically, the ability of the measure to examine different areas of functioning was enticing to those interested in a comprehensive yet efficient assessment of a large group of individuals. For example, as World War II ensued, military psychologists were again concerned with evaluating various dimensions of psychological functioning with intent to identify individuals best fit for different military positions.

Although popular, the BPI had several significant limitations (Super, 1942). Specifically, validity tests revealed that despite claims that the test measures multiple domains, normative

groups did not reliably provide elevated reports on domains reflective of their specific psychological statuses. For example, Greene (2011) indicated that only 39% of patients deemed “neurotic” scored above the 90th percentile on the BPI domain neuroticism. Gabby and Zicker (2008) argued that a potential reason that the BPI has low validity is due to its focus on negative or maladaptive components of the different domains. In transitioning the use of the BPI to an industry setting, it is conceivable that many test takers would have adaptive features of the domains assessed, such as alertness and conscientiousness. However, if the measure does not adequately capture adaptive components of the measure, it would be difficult to adequately measure the constructs in that setting. Ultimately, the BPI appeared to fail to evaluate across the entire spectrum of each domain.

Additionally, Greene (2011) stated that early personality measures, such as the BPI and the Woodsworth Personal Data Sheet were constructed based on a logical approach to item development. Whereas popular approaches to scale development often include an empirical analysis to develop specific items followed by a factor analysis to determine the extent to which items load onto predicted constructs (Spearman, 1904), early personality researchers more heavily relied upon clinical judgment and past experiences to determine whether items were included on assessment measures. Despite criticisms, researchers have since touted the utility and validity of rationale-based approaches to scale item development (Butcher et al., 1990; Wiggins, 1973).

The Humm-Wadsworth Temperament Scale (HWTS; Humm & Wadsworth, 1934) appeared to resolve some of the issues identified in other early personality measures, which have influenced the development of modern assessment measures. The HWTS incorporated a personality theory asserting that maladaptive traits observed in individuals with psychological

pathology or related problems are variations of traits that all people possess. That is somewhat in line with contemporary theories of personality that argue personality traits are continuous constructs, and all individuals register on a spectrum of those traits. In contrast, historical dichotomous theories of personality assert that pathology involves a series of unique traits and behaviors that are entirely distinct from non-pathology. The HWTS consists of seven continuous components of temperament (Humm & Wadsworth, 1934). The Rosanoff theory posits that although those components (normal, antisocial, cycloid manic, cycloid depressed, schizoid autistic, schizoid paranoid, and Epiloid) quantitatively differ between individuals, all individuals register somewhere on each continuum (Rosanoff, 1920). The continuous structure of the HWTS made the measure more broadly applicable to various settings than previous measures. Specifically, the HWTS was ambitiously marketed to and utilized within industrial settings (Gabby & Zicker, 2008).

Overall, the historical progression of objective psychological assessment measures has likely influenced the development of early and more contemporary adaptations of personality assessments used today. Criticisms of early measures include rational approaches to item development, one-dimensional frameworks, dichotomous domain structure, and lack of validity indicators. Despite criticisms, researchers have since utilized some early methodological practices (i.e., rational item selection), as they have proved useful in generating assessment items. The development of the HWTS advanced the development of empirically derived item selection and ultimately informed about development procedures for future and more advanced personality assessment measures. Although early personality measures suffered from methodological flaws that ultimately contributed to reliability and validity problems, they each appear to serve an important role in the foundation of modern assessment measures. Elements of

early assessments will likely be visible in new adaptations of personality measures for the foreseeable future.

In light of advancements to personality assessment, Hathaway and McKinley (1940) sought to compose a new assessment inventory that could build upon the strengths of its predecessors, while also broadening the context of its utility. Hathaway and McKinley (1940) compiled over 1,000 statements to be used as assessment items, consulting their vast clinical experience as well as their access to psychological resources (Greene, 2011). They composed a series of declarative statements to which respondents could simply answer true or false. They utilized an empirical criterion keying method to construct the test. Specifically, researchers first identified known groups of individuals with varying types of psychopathology and personality characteristics. Those groups, called criterion groups, were utilized to establish data informing researchers how respondents would be likely to respond if they demonstrated similar characteristics to the criterion groups. If respondents responded similarly to individuals with known psychopathology, the test could serve as a potential indicator for the presence of that same psychopathology. This empirical criterion keying method addresses problems of external validity of earlier measures. Specifically, it ensures that respondent data are compared with known normative data rather than assumptions about how certain groups would respond to different questions. They eliminated items with redundant content and that appeared insignificant. Greene (2011) stated that though they did not have a systematic approach for eliminating seemingly insignificant items, Hathaway and McKinley (1940) argued that their empirical method for generating items allowed for that flexibility. Ultimately, they reduced their pool to 504 items, which they used to empirically construct multiple quantitative scales specific to types of psychopathology. They examined differences in responses on each item between

criterion and control groups. Observed differences indicated whether a specific item could be considered an indicator of a specific form of psychopathology. The measure they named the Minnesota Multiphasic Personality Inventory (MMPI) was released in 1943 (see Hathaway & McKinley, 1943). By 1946, ten scales designed to measure different dimensions of psychopathology were included on the MMPI (Greene, 2011).

Given the multidimensional nature of the MMPI, developers sought to provide an interpretive output that was relatively straightforward. Since each scale was uniquely composed of different quantities of items with varying content, a profile that displayed raw score elevations across different domains would have little interpretive value. Instead, the MMPI and its successors utilize standard T-scores, which demonstrates both how one score relates to a normative group and how a score on one scale compares to a score on another (Butcher et al., 1989).

MMPI Validity Scales

Although the MMPI gained popularity as an empirically derived measure of psychological functioning, the development of scales designed to evaluate response styles is perhaps one of the tests most revered accolades. Given the objective nature of the MMPI, it was important for the developers to consider the role of test-taker bias in their ultimate performance. External variables can motivate test-takers to respond such that they are likely to obtain profiles that allow them to achieve some external goal. In addition, when test-takers understand the nature of test items, they are more likely able to influence their responses to ultimately affect their desired assessment outcome. Ultimately, these problems can impact the face validity of an assessment measure. Specifically, given the goal of the MMPI was and is to provide an accurate reflection of an individual's psychosocial functioning, motivating factors may impair that profile

and thus, contaminate the results. Scholars interested in assessment validation have long considered protective and risk factors that can affect validity.

In contrast, performance-based measures of assessment have been somewhat successful in mitigating the impact of assessment response bias. The Rorschach, for example, provides information about psychological functioning similarly to the MMPI. However, assessment information is obtained through an examination of how evaluatees interpret an inkblot image. The interpretive quality of individual responses is not readily apparent to most evaluatees, which somewhat protects the measure from intentional response biases (Huprich & Ganellen, 2006). Conversely, objective measures of assessment, such as the MMPI, require that evaluatees consider whether relatively straightforward statements about psychological functioning applies to them. As such, evaluatees who are motivated to present themselves in a certain way in response to internal or external influences could more easily identify how best to respond on objective measures, such that their performance reflects a desired outcome rather than their actual levels of functioning. For example, were an examinee interested in receiving benefits associated with a psychological disorder, he or she might then be motivated to endorse MMPI items that reflect disordered psychological functioning. Although the developers of early objective measures largely conceded that response style was an important factor that could impact the face validity of their assessments, they failed to directly address those concerns in the development of their measures. Greene (2011) described that despite developers providing “lip-service” (p. 10) about the need for consideration of response styles, early objective assessment measures did not include specific methods to detect the impact of response bias on profile validity. The developers of the MMPI sought to address that very limitation of early assessment measures.

As the first edition of the MMPI became available, researchers and clinicians stressed the impact of content non-responsiveness. Several scales were composed of questions specifically designed to evaluate different types of response styles. The Cannot Say (CNS) scale involves a raw score of the number of items omitted or answered both “true” and “false.” The CNS scale is not a series of questions, but rather a count of the number of questions that could not be considered during scoring. When scales are constructed, the reliability and validity of those scales is predicated on the content of each of its individual components. Thus, the omission of enough items can effectively change a MMPI profile.

To address the potential impact of test-taking response styles, Meehl and Hathaway (1946) identified both defensiveness as well as excessive reporting as two distinguishable response styles that could impact the validity of a MMPI profile. They assert that defensiveness occurs when test-takers are motivated to underreport the extent of psychological problems they experience. In other words, they present themselves with an overly positive valance. Excessive overreporting, on the other hand, occurs when test-takers over represent the extent to which they experience psychological dysfunction and/or distress. To develop test items that could reliably evaluate response styles that threaten assessment validity, Meehl and Hathaway (1946) compiled items infrequently endorsed by different normative groups that reflect both over exaggeration and under reporting. They discuss that test-takers inclined to over represent their symptoms will not only affirm the presence of psychopathology seen in clinical populations, but they will also often affirm items indicating psychological dysfunction that are not commonly seen in clinical populations. The Infrequency scale (F scale) on the MMPI consists of 64 items that highlight unfavorable characteristics or experiences rarely endorsed by the MMPI normative population.

As such, individuals who endorse a high number of F scale items are likely over representing the extent of their psychological distress or dysfunction.

Meehl and Hathaway (1946) also utilized infrequently endorsed items to identify the presence of overly self-favorable responding. Specifically, they argued that the design of items consisting of characteristics that were considered highly desirable yet infrequently endorsed by populations without an incentive to over report could reflect a scale that identifies individuals who underreport the presence of genuine problems or pathology. Several scales were initially developed to address the issue of the self-favorable response styles on the MMPI. Meehl and Hathaway (1946) identified that test takers could provide self-favorable impressions in several ways. For example, they argued that some individuals were likely to endorse highly virtuous characteristics that were uncommon within the normal population. As such, they composed the Lie scale (L scale), which consisted of 15 items addressing the presence of such uncommon virtues. Additionally, Meehl and Hathaway (1946) identified a need to distinguish individuals who are interested in self preservation related to their level of functioning. They noted that whereas some individuals inaccurately report the extent to which they exhibit virtuous characteristics as measured by the L scale, some individuals instead underreport the presence of genuine psychopathology or other abnormal functioning. There are likely many internal and external variables that can influence the accuracy of reporting symptoms of psychopathology, many of which will be discussed in forthcoming sections. The Correction scale (K scale) was added, consisting of 30 items that address symptoms of psychopathology that are infrequently endorsed by individuals with genuine psychopathology. Elevations on K suggest an individual is displaying a defensive quality and possibly underreporting the presence of genuine

psychopathology. K can be influenced by a desire to appear healthy or free of psychological impairment.

Developers of the original MMPI were attuned to the drawbacks and advantages of previously designed objective assessment measures. They utilized the broad utility of dimensional constructs on which all individuals can be measured. Additionally, they identified the practicality and utility of a multidimensional measure that evaluates various aspects of psychological functioning. The validity scales that were incorporated with the innovative MMPI added an important component of psychological assessment that likely has influenced the longevity of the measure in the decades following its initial conceptualization. For the first time, they allowed evaluators to identify whether specific test taking attitudes influenced ultimate profile reports. The development and advancement of the validity scales have likely contributed to the utility of the MMPI in multiple contexts, particularly those where there is concern about response bias.

MMPI-2 Validity Scales

Following several decades of use, researchers began to express the need for revisions as well as a restandardization of the MMPI normative sample. The original normative sample consisted of men and women accompanying relatives to the University of Minnesota Hospital (Pancoast & Archer, 1989). Among other things, the original MMPI was criticized for its use of a normative sample that was relatively non-representative of the U.S. population. Calls were made to develop a normative reference group that better represented gender and ethnic diversity in the broader population. As is often the case with well-established tools in various contexts, scholars were conflicted about the prospect of revisions to the original MMPI. Butcher (2000) indicated the importance of such dissent in the development of the MMPI-2. He argued that

careful consideration should be given to aspects that are in need of change and aspects that should be held constant across revisions. Butcher (2000) suggested that a major advantage of the MMPI-2 revision was the careful consideration given to the preservation of valid components of the original MMPI. Given the extensive research used to validate the original validity scales (for meta-analytic reviews see Baer et al., 1992; Berry et al., 1991), the composition of the existing MMPI validity scales was largely maintained throughout the development of the MMPI-2. Specifically, scales CNS, L, and K went completely unchanged on the MMPI-2, whereas the only changes to the F scale included the elimination of four “problematic” items (Wygant, 2018, p. 259).

A primary goal of the MMPI-2 workgroup was to focus on the development of new comprehensive norms. Pancoast and Archer (1989) composed a meta-analytic study examining “normal” adults in comparison with MMPI normative data. They found significant differences between the groups, suggesting that the original normative group no longer represented a valid comparison. The MMPI-2 normative sample represented a more diverse sample of the population, including individuals with various education levels, ethnic backgrounds, and occupational backgrounds (Greene, 2011).

Importantly, the development of the MMPI-2 also included new validity scales. The Back Infrequency (Fb) scale consisted of 40 items that evaluated infrequent responding specifically toward the end of the assessment. In addition, item content on the Fb scale varies from the F scale. Specifically, an individual who exhibits characteristics of depression, suicidal ideation/behavior, and hopelessness, but who is not psychotic, would be more likely to demonstrate an elevation on the Fb scale than the F scale. The Variable Response Inconsistency (VRIN) and True Response Inconsistency (TRIN) scales consisted of item pairs designed to

identify the presence of random and fixed response styles respectively. Additionally, multiple supplementary validity scales were also developed at the conceptualization of the MMPI-2 (Wygant et al., 2018). For example, the Symptom Validity Scale (FBS) was developed as a validity scale specifically attune to exaggerated cognitive or somatic symptoms rather than psychotic symptoms (Lees-Haley et al., 1991). Additionally, in response to criticisms indicating the F scale had a high false-positive rate, Arbisi and Ben-Porath (1995) developed the Infrequency Psychopathology (Fp) scale, which consisted of a subset of F scale items and was more effective at distinguishing genuine from feigned psychopathology (Rogers et al., 2003). The Fp scale was normed using an inpatient sample and is less sensitive to extreme response tendency, like as F and Fb. Arbisi and Ben-Porath (1998) found that the Fp scale was more effective in distinguishing feigned psychopathology than the F scale.

MMPI-2-RF Revision and Validity Scales

Criticisms and the need for updates prompted an updated version of the MMPI-2. Critics argued that the MMPI-2 consisted of scales that frequently overlapped and did not adequately distinguish psychological characteristics and syndromes. In addition, critics argued that pervasive characteristics of “demoralization” appeared to span across domains, contributing to difficulty distinguishing between the different domains. The MMPI-2 clinical scales were criticized for their overly heterogeneous content and high intercorrelations between scales. Further, a longstanding criticism of the MMPI and MMPI-2 was that they included too many items, which contributed to test-taker fatigue and long administration timeframes. The MMPI-2-Revised Form (RF) was designed to address the different concerns presented throughout the literature. Specifically, it was designed to be more efficient and able to distinguish each clinical subscale from the associated subjective distress (Greene, 2011).

In contrast to prior editions, the MMPI-2-RF utilized a factor analytic approach to develop the RC scales, ultimately to reduce item overlap between scales that resulted from the criterion keying approach. Factor analysis removed items that previously loaded onto multiple clinical scales. Following factor analyses, items that highly loaded on a “demoralization” variable were extracted from their original scales. Those items were theoretically related and comprised the 24-item “demoralization” scale in the MMPI-2-RF. Following that, remaining items for each clinical scale theoretically represented the substantive core of each scale (Ben-Porath & Tellegen, 2008). From that, demoralization and 11 other distinctive *seed* scales became apparent. The 12 scales were correlated with the items from the MMPI-2 and considered for item content, ultimately producing the RC scales used on the MMPI-2-RF. In addition, the revised form consists of 338 items in contrast to the 567 items on the MMPI-2, addressing concerns about test-taker fatigue.

As with the previous revision, the validity scales were largely retained on the MMPI-2-RF, with a reduction in item quantity and the addition of several new scales as the extent of the revisions. The MMPI-2-RF includes both non-content-based validity scales, which analyze item consistency and item completion, as well as content-based validity scales, which examine the impact of specific response biases.

Cannot Say (CNS) Scale. As with the MMPI-2, the first step in determining the validity of the MMPI-2-RF involves examining the number of items that cannot be scored. An examination of raw data collected from normative, clinical, and personnel samples suggest that even a single omission on the MMPI-2-RF is rare (Greene, 2011). Research has examined the extent to which omitted items can impact clinical scale profiles. Berry et al. (1997) found that just 10 omissions contributed to significant profile distortions in 25% of their MMPI-2 sample.

Given that the MMPI-2-RF consists of fewer items, omissions have a greater impact on test validity.

Variable Response Inconsistency-Revised (VRIN-r). To improve the validity of the VRIN-r scale from its predecessor, items from the MMPI-2 that were highly correlated ($r > .90$) were included. Ultimately, 53 item pairs were selected for the VRIN-r scale, which demonstrate content consistency. Whereas some items are reverse scored, others are scored in the same direction as their counterparts. A *T* score of >80 on the VRIN-r invalidates a protocol (Ben-Porath & Tellegen, 2008).

True Response Inconsistency-Revised (TRIN-r). Similar to the VRIN-r, Ben-Porath and Tellegen (2008) compiled 26 item pairs that are highly correlated ($> .90$) when composing the TRIN-r. In contrast to the VRIN-r, item pairs not only measure content consistency, but also scoring consistency. Specifically, items are considered inconsistent when both are scored either “true” or “false.” That allows for the test to determine whether test-takers engage in regimented responding regardless of item content. A *T* score of > 80 or < 20 on the TRIN-r invalidates a protocol (Ben-Porath & Tellegen, 2008). A high score suggests a tendency to score “true” whereas a low score suggests a tendency to score “false.”

Infrequency-Revised (F-r). The F-r consists of 32 items that were infrequently endorsed within the normative sample (Ben-Porath, 2012). Though it is sensitive to overreporting as well as inconsistent responding, critics have argued that individuals with genuine psychopathology tend to also obtain elevations on the F-r scale. Additionally, Greene (2011) suggested that the items included in the scale are susceptible to test-taker response bias due to the obvious nature of the item content. Despite its drawbacks, the F-r is reliably correlated with prior infrequency

scales and can serve as a valid indicator of psychological distress and can serve as an indicator of overreporting.

Infrequency-Psychopathology (Fp-r). The Fp-r consists of 21 items that were infrequently endorsed by a psychiatric normative sample. Several items were removed from the previous edition due to overlap with other scales. The Fp-r is correlated with its predecessor, which was regarded as the most predictive validity scale used to predict over reporting (Rogers et al., 2003).

Infrequent Somatic Responses (Fs-r). The Fs-r normative sample involved patients with known medical diseases. The scale includes 16 items involving somatic symptoms infrequently endorsed by those medical patients. The items were derived from an original list of 120 items. The final 16 include physical ailments and dysfunction rarely endorsed by medical patients. Greene (2011) suggested that a *T* score above 74 on the Fs-r indicates that a respondent is endorsing a high number of atypical ailments not usually reported by medical patients.

Symptom Validity Scale (FBS-r). The FBS-r was largely derived from the original Faking Bad Scale (Lees-Haley et al., 1991), including 30 of the original 43 items. Although the FBS label was retained, the authors renamed the scale the Symptom Validity Scale to reduce the pejorative nature of the title. The criterion sample includes personal injury plaintiffs with legitimate somatic symptoms as well as medical patients with known medical diseases. Greene (2011) highlighted that 95% of personal injury litigants from two separate samples did not provide elevated responses on the FBS-r, suggesting that personal injury litigants largely display genuine somatic symptoms. A *T* score of 65 or greater suggests the presence of atypical or unusual somatic experiences not typically reported by medical patients or personal injury litigants.

Response Bias Scale (RBS). The RBS was introduced as a 28-item scale that was empirically derived from relevant cognitive assessment validity measures (Gervais et al., 2007). Gervais et al. (2007) found that performance on cognitive validity tests was strongly correlated with performance on the RBS scales, particularly in forensic and disability samples. As such, the RBS scale has been utilized to distinguish between genuine and biased reports of cognitive dysfunction.

Uncommon Virtues (L-r). The L-r scale is used to identify the presence of self-favorable response bias on the MMPI-2-RF. The scale consists of 11 of items from the original L scale and 3 additional items. As such, the L-r scale is correlated with earlier editions and consists of items involving minor personal faults that even the most virtuous individuals are willing to acknowledge when responding truthfully. A *T* score below 44 suggests an effort to provide an extremely unfavorable picture of oneself. In contrast, A *T* score above 65 suggests possible lack of intrapersonal insight or an attempt to provide an overly virtuous self-evaluation.

Adjustment Validity Scale (K-r). Of the original 16 items on the K-scale, 14 were retained for the K-r scale. In contrast the L-r scale, the K-r scale focuses more on respondents' level of psychological functioning and the presence of psychopathology. A *T* score below 35 suggests extreme distress due to psychopathology and little autonomy to make significant changes, whereas a *T* score above 65 suggests either a lack of insight into psychological weaknesses or a reluctance to report any form of genuine psychopathology.

MMPI-2-RF-EX & MMPI-3

The development of the MMPI-3 sought to improve upon the strengths of previous editions of the MMPI. The MMPI-2-RF-EX was developed as a research instrument to examine the items to be used on the MMPI-3. The MMPI-2-RF-EX consists of 433 questions. It contains

both the items included on the MMPI-3 as well as additional research items. Analyses of MMPI-2-RF-EX and MMPI-3 scores suggest good overall comparative validity (Hall et al., 2020). The MMPI-3 includes 335 items, which is comparable to the number of items on the MMPI-2-RF. A notable strength of the MMPI-2-RF was the reduction in the number of items included in comparison to its predecessor. Having fewer items contributes to reduced administration time as well as reduced test-taker fatigue. MMPI-3 validity scale development was heavily influenced by the composition of the MMPI-2-RF validity scales. The revised validity scales offer a refined examination of response bias that is applicable in multiple contexts. There is extant research on validity scales, which informed both the refinement of existing scales, as well as the development of new scales.

Non-Content-based Validity Scales. As with the MMPI-2-RF, VRIN and TRIN on the MMPI-3 were developed through a composition of item pairs, the answers to which determine variable or fixed responding. VRIN is composed of 53 composite scores whereas TRIN is composed of 33 composite scores. New to the MMPI-3, the Combined Response Inconsistency (CRIN) consists of 86 inconsistent response composites that are derived from VRIN and TRIN. The CRIN was developed as a broader indicator of inconsistent responding.

Overreporting Validity Scales. Although the F-r scale in the previous edition set a criterion suggesting that no more than 10% of the normative population should endorse F-r items, a less strict criterion was used for the MMPI-3, such that a sufficient number of response indices could be identified. Specifically, a 15% criterion was used for the MMPI-3 F scale. Twenty-six items of the MMPI-2-RF compose the 35-item MMPI-3 F scale. Given that previous research has found elevations in scale F within honest responding forensic and psychiatric

inpatient samples, it is likely that that finding will become even more present when a stricter criterion is utilized.

On the MMPI-2-RF, the Fp-r scale utilized a 20% criterion value for item inclusion, which was retained on the MMPI-3. With the new normative sample of mental health patients, 17 items were retained for the MMPI-3 and the others were replaced with MMPI-3 items that met the criterion. The MMPI-2-RF sample utilized for the Fs scale consisted of medical patients. The scale consisted of items that were endorsed by fewer than 25% of the sample. Of the 16 items on the MMPI-3 Fs scale, 13 were retained from the Fs-r scale. Those that no longer met the criterion were replaced. The FBS (30 items) and RBS (28 items) went unchanged on the MMPI-3.

Underreporting Validity Scales. The MMPI-3 Uncommon Virtues (L) and Adjustment Validity (K) scales were updated to include items that met the criteria, but remained largely similar to their predecessors on the MMPI-2-RF.

The MMPI-3 yielded somewhat modest changes to the validity scales, which reflect their longstanding stability. An examination of the advancement of the validity scales suggests that the application of the scales in multiple contexts and under various contextual conditions has contributed to a more nuanced understanding of each and has led to the development of more specified scales. Researchers have learned more about how unique populations under different conditions uniquely respond on validity indices, providing information about the scale itself as well as the population examined (Ray, 2017). Whereas the F scale was once thought to capture overreporting in its entirety, differential performance in various contexts has contributed to clarity of the specific utility of that scale as well of the development of new scales to examine subgroups of sample populations. With the release of the MMPI-3, it will be important to

continue to investigate how the validity scales perform with various populations under different contextual circumstances to best develop normative expectations.

Forensic Applications and Legal Standing of the MMPI

Psychological testing has been a cornerstone of psychologists' contributions across various contexts. Specifically, psychological testing has aided in important clinical decision making in contexts including, psychiatric, outpatient clinical, and forensic settings (Groth-Marnat, 2009). In fact, psychological testing is often a distinguishing factor that identifies the unique role of psychologists (Melton et al., 2018). The addition of empirically based assessment measures to a seemingly otherwise subjective evaluation can contribute an added element of objectivity. Although testing is not always necessary in psychological evaluations, psychologists and legal professionals have relied on various assessment measures in forensic contexts for a variety of purposes. In forensic contexts, referring to statistical data derived from empirically scrutinized assessment measures can not only influence decision makers' decisions, but also add validity to psychological evaluations. Psychologists asked to provide opinions about psychological processes can rely on mathematical probability as one piece of evidence to support their conclusions in the face of strict legal scrutiny. MMPI instruments have been commonly utilized in various forensic examinations likely due to the broad nature of the measures as well as their adherence to legal standards regarding the application of science to law.

Daubert v. Merrell Dow Pharmaceuticals (1993) was a landmark case that ultimately shaped how scientific testimony and evidence is considered in legal contexts. As such, the Daubert Standard asserts that the five tenets that may be considered in determining whether the methodology is valid are (a) whether the theory or technique in question can be and has been tested, (b) whether it has been subjected to peer review and publication, (c) its known or

potential error rate, (d) the existence and maintenance of standards controlling its operation, and (e) whether it has attracted widespread acceptance within a relevant scientific community (Melton, 2018). Given the historical legal scrutiny of the use of scientific opinion in court, psychological assessment measures applied to forensic contexts arguably endure an added element of scrutiny during development, validation, and implementation. Indeed, Vallabhajosula and van Gorp (2001) identified that other commonly used forensic assessments adhere to the Daubert Standards. Specifically, they examined the Rey Fifteen-Item Test (FIT; Rey, 1964), the Test of Memory Malingering (TOMM; Rees et al., 1998), and the Validity Indicator Profile (VIP; Frederick & Crosby, 2000).

The MMPI and its revised counterparts have undergone scrutiny as to whether the measures adhere to the Daubert Standards. When surveyed, 94% of forensic psychologists opined that the MMPI-2 met the Daubert Standards (Bow et al., 2010). Specifically, when surveyed, practitioners preferred the use of the MMPI-2-RF to the MMPI-2 due to its brevity, more narrow focus of validity indices, and empirical integrity. In addition to survey data, professional organizations provide workshops and continuing education credits related to the MMPI-2-RF and entire graduate courses in psychology are devoted to the practical use of the MMPI-2-RF (Ben-Porath, 2012b). In consideration of the Daubert standard, which assumes the widespread acceptance of a measure, MMPI measures appear to have been well accepted in multiple contexts. After all, research suggests that as of 2014, the MMPI-2 was the measure most utilized in forensic evaluations (Niel & Grisso, 2014). Sufficient data had not yet been compiled regarding the use of the MMPI-2-RF. The authors anticipated that the MMPI-2-RF would soon eclipse the MMPI-2 as the most frequently used assessment measure in forensic evaluations. Ben-Porath (2012b) also cited the wealth of literature accumulated on MMPI instruments as

evidence that the measure has held up to scientific scrutiny. Although the MMPI-2-RF has been defended regarding its validity in forensic contexts (see Ben-Porath, 2012b), others have indicated that the MMPI-2-RF is vulnerable to challenges of the Daubert Standard. Graham (2012) argued that the MMPI-2-RF is inherently distinct from the MMPI-2 due to differences between the measures and, therefore, should not be defended using norms and research established using the MMPI-2. However, researchers have defended the application of MMPI-2 research to MMPI-2-RF data. Whereas the development of the MMPI-2 included a new normative data sample, the MMPI-2-RF utilized the normative sample used for the MMPI-2. (Ben-Porath & Tellegen, 2008). MMPI-2-RF developers compared 561 MMPI-2-RF administrations with the normative sample from the MMPI-2. They found modest differences in four scales (Fs, L, DSF, COG). Overall, the MMPI-2-RF sample data appeared to match on well with normative data. Given that the MMPI-2-RF was developed using MMPI-2 items and normative samples, it is likely that research using the MMPI-2 contributes to the validity of the MMPI-2-RF.

The Daubert Standard has been used to monitor the use of “scientific evidence” in legal context. Though there is some dissent about the admissibility, legal psychologists seem to have largely accepted MMPI instruments as valid indicators of response style and psychological functioning. The research identified in previous sections as well as an analysis of the Daubert Standards suggest that MMPI instruments may be effective in aiding in clinical decisions made by forensic psychologists. Research suggests that MMPI instruments have been effective in distinguishing between subgroups of forensic populations who have been assigned unique forensic opinions.

Regarding the use of the MMPI-3 and the present research, there is cause for argument about whether the new measure continues to meet Daubert standards, despite the inclusion of a new normative sample in addition to item updates. Indeed, the MMPI-3 has yet to be “generally accepted by the scientific community” given the measure has only recently been released. However, consideration of the tenets of the Daubert standard suggest the MMPI-3 has been and can be empirically tested, is based on research evidence, and is standardized, indicating the measure meets at least the minimum standards to be utilized in forensic applications. The historical scrutiny placed on psychological assessment measures in legal settings and release of the MMPI-3 underscores the importance of developing a foundational research base that can provide clarity regarding group performance on the measure. The present study aims to contribute to that body of research.

Forensic Populations

Forensic contexts often involve high-stakes evaluations and assessments that examine different types of individuals. Forensic institutions are often designated to provide evaluation and intervention services to several groups of individuals with mental health concerns and who are also involved in the legal system. Relevant groups include individuals suspected to be or found incompetent to stand trial (IST) or individuals suspected to be or found not guilty by reason of insanity (NGRI). Although individuals in such groups often have similarities regarding legal infractions and restrictions, as well as clinical symptoms, there are unique elements of each group that are important to distinguish.

Incompetent to Stand Trial (IST)

The necessity of a defendant to aid in his or her own litigation through factual understanding and rational decision-making is a foundational cornerstone of the United States

judicial system. When the ability to rationally understand and contribute to one's defense is impeded due to a diminished mental capacity, continued litigation would infringe on defendant rights under the United States constitution (Melton et al., 2018). IST defendants are unique from other defendants in that they experience a mental condition that inhibits their rational and/or factual understanding of the charges against them as well as their sufficient and present ability to assist and consult with their attorney in preparing a legal defense. Although their mental conditions during legal proceedings *can* and often do reflect active mental illnesses experienced during alleged offenses, mental illnesses that affect competency to stand trial (CST) alone do not necessarily substantiate an insanity defense. Whereas an insanity defense requires a retroactive examination of a prior mental state that impacted the ability to distinguish right from wrong at the time of an alleged offense, IST status reflects current mental status and how that status impacts their ability to understand basic legal information and make decisions, along with an attorney, regarding how defendants wish to plead and proceed through trial. Individuals undergoing IST restoration often receive psychiatric treatment, which can have implications on psychological testing. Defense attorneys who suspect a client is unfit to proceed often request an evaluation for competency, which can subsequently lead to restorative treatment. There has been considerable research examining commonalities between IST defendants.

An early study examining CST evaluations suggests that disorganized speech was the strongest predictor of referrals by attorneys for competency evaluations (see Berman & Osborne, 1987). That information is understandable given the importance of communication between attorneys and their clients. Additionally, in a recent meta-analysis, Piralli et al. (2011) found that defendants with psychotic symptoms were more likely to be opined IST than defendants without intrusive psychotic symptoms. Further, they found that IST defendants often had extensive

mental health and/or criminal histories to a greater extent than other forensic populations. Although inpatient IST defendants and inpatient psychiatric (non-IST) adjudicates demonstrated similar prevalence rates of psychotic disorder diagnoses, inpatient IST defendants reported significantly higher *symptoms* of psychosis. Further, Riley (1998) found that competency was negatively related to psychotic symptoms. The results suggest that individuals with psychotic disorders are at risk for problems related to competency. However, the results also suggest that current symptomology, rather than historical diagnoses alone, are most predictive of incompetency.

Viljoen et al. (2002) further examined the relation between psychosis and IST status. They found that 60% of their sample of patients with a diagnosis of a psychotic disorder did not demonstrate legal knowledge impairment, ultimately suggesting that not all defendants with a psychotic disorder experience active psychosis that impairs their legal competency. They also found that individuals with schizophrenia demonstrated greater impairment in legal knowledge than individuals with other disorders that can include symptoms of psychosis. They found weaker correlations between other mental illness (e.g., bipolar, depression, cognitive dysfunction) and CST. They cite the greater likelihood for symptoms of psychosis to impair CST factors as a potential rationale for their findings. Given that schizophrenia often reflects more persistent and severe psychotic symptomology than some other psychotic disorders, that finding is understandable. Indeed, more severe symptomatology has shown to be highly correlated with impaired legal knowledge (Hoge et al., 1997; James et al., 2001).

Intellectual and educational ability have also been found to impact CST. Viljoen et al. (2002) found that whereas IQ was not a significant predictor of how well defendants understood or appreciated the consequences of the proceedings, IQ significantly predicted defendant abilities

to understand the nature and object of proceedings as well as interrogation warnings. Individuals evaluated for CST frequently have below-average education levels, impacting the ability to understand difficult legal concepts (Morse & Morse, 1980). IST evaluatees and inpatients often display impairments related to low IQ and/or psychiatric symptomology (Nicholson & Kugler, 1991). Morse and Morse (1980) identified that individuals evaluated for CST were found to have significantly lower educational ability than other defendants, contributing to greater difficulty understanding complex information. In addition, Hoge et al. (1997) and Ustad et al. (1996) found that defendants evaluated for competency often have significantly lower IQ scores, particularly in verbal reasoning than other defendants. Further, they found that incompetent defendants have more impaired cognitive functioning when compared to defendants found competent. More recently, Matlasz et al. (2017) found that IST defendants were more likely to have lower intelligence scores than non-IST defendants as measured by the WAIS-IV.

Overall, research suggests that the mere presence of a psychological disorder is not indicative of an individual's CST. Rather a combination of symptomology, educational ability, and intelligence all play a role in CST. The highlighted research demonstrates further evidence that defendants evaluated for CST are a unique subpopulation within a broader forensic psychiatric sample.

Not Guilty by Reason of Insanity (NGRI)

NGRI is a plea option for criminal defendants and a legal decision determined by a judge and jury. Although each state maintains their own statutes related to legal sanity, most states follow the M'Naghten precedent for legal insanity defenses. Under the M'Naghten rules, a defendant is not guilty by reason of insanity if, at the time of the alleged criminal act, the defendant was so deranged that he or she did not know the nature and quality of his or

her actions, or if she knew the nature and quality of her actions, she knew the nature and quality of her actions, she was so deranged that she did not know that what she was doing was wrong. (Queen v. M'Naghten, 1843, p. 1)

NGRI defendants acknowledge that they engaged in illegal behavior but argue that they suffered from such a defect of reason due to a mental illness or impairment, that they were unable to understand that their behavior was wrong at the time of the alleged offense. In contrast to individuals opined IST, NGRI adjudicates have completed the legal process as it pertains to their related offense. NGRI evaluatees, on the other hand, have yet to receive an ultimate judicial decision regarding their plea and are actively engaged in the legal process. Although contextual phenomena are likely to impact how each group of NGRI defendants present themselves, NGRI adjudicates compose a distinct group of individuals who have unique clinical presentations. Whereas CST is one cornerstone of the American judicial system, so too is the notion that the law should punish only those who “choose to commit crimes for rational reasons and of their own free will, and are therefore deserving of punishment” (Melton et al., 2018, p. 199). When mental illness is a consideration, authorities are tasked to determine whether the crime in question was a result of that mental illness and whether it sufficiently impaired the defendant’s ability to recognize the illegality of their behavior. *Durham v. United States* (1954) established that tests for insanity should identify the presence of mental disease or defect as the cause for criminal behavior in NGRI cases. The Durham Rule considers “disease” synonymous with “mental illness” and “defect” synonymous with “intellectual disability.”

Scholars have examined the frequency of different psychological diagnoses and syndromes that result in successful NGRI defenses. Research suggests that psychotic symptoms are most commonly observed in successful NGRI defenses, whereas intellectual disability, and

mood regulation are only somewhat frequently used in successful NGRI defenses (Melton, 2018; Warren et al., 1991). Melton (2018) highlighted that research examining psychologist opinions and judicial decisions regarding NGRI reflect similar prevalence rates of specific impairments. Specifically, psychotic disorders are most likely to lead to NGRI psychologist opinions and ultimate adjudication. Additionally, intellectual and mood impairments are somewhat likely to be present in individuals opined NGRI. Importantly, Melton (2018) highlighted that whereas schizophrenia has an overall prevalence rate of about 1%, NGRI defenses are only proposed in about 1% of legal cases and are even less frequently successful. The NGRI defense poses a high bar for a defendant to qualify.

Melton et al. (2018) examined several studies that identified common characteristics across NGRI adjudicates. They found that, in contrast to IST defendants, patients most often found NGRI are more likely to have a scant history of prior psychiatric hospitalizations. That suggests a notable difference between IST and NGRI defendants. Research suggests that a majority of IST patients have extensive psychiatric and legal histories. It is true that some NGRI adjudicates are initially found IST and/or have lengthy psychiatric histories. However, the aforementioned research suggests that NGRI evaluatees and adjudicates may have distinct psychological histories on average. Melton et al.'s (2018) findings are understandable given that psychotic symptoms frequently do not present until individuals reach their twenties. As such, individuals who engage in legitimate NGRI offenses may have done so at the onset of their symptoms. The results suggest that NGRI adjudicates may not consistently have as lengthy of a psychiatric or criminal history as IST defendants, whereas IST status is more likely to be impacted by factors other than symptoms of a severe mental illness. Melton et al.'s (2018) results also suggest that few successful NGRI defenses do not include psychosis. Consistent with prior

research, the majority of NGRI defendants and adjudicates experience psychotic symptoms that contributed to their NGRI offenses.

NGRI adjudicates have also been considered in comparison with individuals found criminally responsible. Nestor and Haycock (1997) found that, when compared to individuals deemed criminally responsible, NGRI adjudicates were found to experience more genuine psychotic symptoms during their offenses. That result is unsurprising given that some individuals deemed criminally responsible are likely to feign elements of their clinical presentations during NGRI evaluations. The authors also suggest that psychosis is a symptom frequently seen amongst NGRI adjudicates, an assertion that has been maintained throughout the literature. Additionally, the authors determined that NGRI adjudicates did not differ from individuals found criminally responsible on measures of psychological intelligence. The results suggest that NGRI adjudicates do not necessarily experience intellectual impairment, which contrasts with research on IST defendants who more often experience intellectual impairment as a contributing factor to their IST status. Although NGRI adjudicates can experience intellectual impairments, the research suggests that NGRI adjudicates commonly suffer from impairments that leave neuropsychological abilities intact.

Although disorders such as developmental/intellectual disability have been used in successful NGRI defenses, the literature on NGRI adjudicates suggests that predominantly psychotic symptoms are used in successful NGRI defenses. Warren et al. (1991) identified that schizophrenia was the most commonly cited diagnosis associated with an insanity defense (28%), whereas intellectual disability was utilized in 11%, and affective disorders were utilized in 15% of insanity defenses. As such, NGRI populations are likely to be composed of individuals who experience delusions or hallucinations that results from psychosis. In addition, the literature

suggests that NGRI adjudicates do not inherently experience intellectual or educational impairments, such as those more frequently observed in IST defendants (Nestor & Haycock, 1997). Lastly, Nestor and Haycock (1997) suggested that NGRI adjudicates are just as likely as not to have no prior psychiatric or legal history. Overall, the literature provides evidence that NGRI, IST, and healthy forensic individuals are distinct samples that each consist of unique characteristics. As such, it may be important for research to consider how each group uniquely approaches psychological assessments. That information could provide examiners with clearer expectations for specific behaviors within different forensic sub-contexts.

Use of MMPI Validity Scales With Inpatient and Forensic Populations

The MMPI and its successors have been used in a variety of contexts (Groth-Marnat, 2009). Given its multidimensional nature, the MMPI has been effective in examining varying problems often seen amongst psychiatric inpatients. As the present study considers the test-taking approaches of inpatient forensic psychiatric patients, it is important to consider how different validity scales have performed uniquely in inpatient as well as forensic populations. A review of the literature will highlight that both the extent to which psychopathology symptoms are managed as well as external pressures can impact validity scale performance.

CST & Non-Content-Based Validity Scales (CNS, VRIN, TRIN). Non-content-based validity scales are useful in detecting omissions and random, fixed, or inconsistent responding that could be due to carelessness, lack of comprehension, or a motivation to distort assessment results. Greene (2007) discussed that non-content-based validity scales are particularly useful because they are the only scales that offer information about the *performance* of a test-taker, in contrast to other validity scales that offer inferences about the foundations for certain self-report profiles. Specifically, non-content-based validity scales measure the extent to which test-takers

actually omit items or engage in fixed or random responding. Other validity scales use probability estimates to generate a likely conclusion about response bias, which is implicated by a greater chance of error. As such, CNS, VRIN, and TRIN have been utilized with inpatient as well as forensic populations.

CNS Scale. Clopton and Nuringer (1977) examined whether distinct groups were likely to display different omission rates on the MMPI using the CNS scale. They found that although psychiatric inpatients, clinical outpatients, and non-clinical job applicants were all likely to have low omission rates, clinical outpatients and psychiatric inpatients were likely to have significantly more omissions than non-clinical participants. They indicated that the presence of psychopathology *can* impact test-takers' ability to effectively attend to or comprehend the test, leading to missed items. Additional research suggests that context can also impact omission rates. Samuel et al. (2007) found that individuals within a forensic context demonstrated higher omission rates than expected. Over 50% of their sample of personal injury litigants omitted at least 11 items. The aforementioned research suggests that psychopathology as well as context might play a role in the extent to which individuals omit items on the MMPI. Research on the CNS in different samples suggests that though omission rates are likely similarly higher among all clinical populations than non-clinical populations, the forensic context might influence even higher omission rates. As a result, individuals with genuine psychopathology who are also within a forensic context might display higher CNS ratings than other populations.

VRIN & TRIN. Other non-content-based validity scales have also been examined in inpatient and forensic settings. Greene (2011) compared the performance of VRIN and TRIN within several different samples (normal, clinical outpatient, inpatient). He concluded that individuals across the groups largely perform similarly on TRIN, about 75% of each ranging in

score from 8 to 10. In contrast, he found that comparative samples performed slightly differently on VRIN. However, although psychiatric inpatients were likely to have higher scores on VRIN than normal individuals, Greene (2011) attributed that finding to the lower education levels observed in the clinical sample. Indeed, Ben-Porath (2012b) suggested that variable responding might occur as a result of intellectual limitations that would be impacted by education level. As such, psychiatric inpatients with low reading abilities are likely to have difficulty interpreting items on the MMPI effectively similarly to normal adults, leading to higher rates of VRIN. Were the education levels consistent across groups, performance on VRIN likely would have been similar.

Himsl et al. (2017) examined psychiatric patient performance on the WRAT-4 (standardized measure of academic achievement) as well response styles on the MMPI-2-RF. They found that elevations in VRIN are inversely related to sentence comprehension level. In other words, psychiatric patients with impaired reading abilities are more likely to variably respond to MMPI-2-RF items than those without impaired reading abilities. The results suggest that it is important to consider sentence comprehension ability when administering the MMPI-2-RF to psychiatric patients. In forensic contexts, patients opined incompetent to stand trial (IST) often have intellectual difficulties that prevent them from attaining legal knowledge (Nestor et al., 1999). Therefore, it is conceivable that individuals who are deemed not competent to aid in their own criminal defenses may also have difficulty responding to MMPI instruments in a consistent manner. Evidence also suggests that presence of active symptomology can impact cognitive function and the ability to perform consistently.

Psychiatric patients also often present with symptoms that impact their daily functionality. Psychotic symptoms can consume cognitive resources and limit attention span.

Carpenter et al. (2000) found that patients with schizophrenia initially displayed cognitive impairment that could impact their ability to make decisions and engage in cognitive tasks. They added, however, that medications as well as psychoeducation regarding symptom management was successful in improving those impairments, such that patients could engage effectively in cognitively stimulating tasks such as the MMPI. Interestingly, there has been little specific focus, in the literature, on the impact of non-stabilized psychiatric illness on reading ability, possibly due to the difficulty of administering an assessment prior to stabilization. The aforementioned research suggests that the presence of non-stabilized symptoms, such as psychotic symptoms, that consume cognitive resources are likely to impair reading comprehension ability to a greater extent than when stabilized. In forensic contexts, patients with psychiatric illnesses are frequently found IST due to impairments in functioning (symptomatology, low IQ), which may impact their performance on psychological assessments. Gu et al. (2017) examined the extent to which forensic inpatients found IST were unable to provide valid MMPI-2-RF profiles. They found that IST inpatients provided significantly greater elevations on VRIN-r and extreme scores on TRIN-r than non-IST patients likely due to cognitive interference resultant from their high symptomatology, or their lack of investment in the task. In addition, they found that in comparison with other studies that utilize forensic samples comprised of both IST and NGRI participants, a significantly larger proportion of their sample provided invalid MMPI-2-RF protocols due to elevations in VRIN-r and TRIN-r. Since Gu et al. (2017) examined IST patients only, it appears likely that individuals who are incompetent will provide higher scores on VRIN and TRIN scales than other forensic populations. Further, researchers have found that scales on other measures of response style that are similar to VRIN and TRIN, which also measure response consistency and bias, were successful in differentiating IST and non-IST defendants (Matlasz et al., 2017).

Though there is research linking IST status to scores on VRIN and TRIN on MMPI instruments, there was no research found examining differences between inpatients undergoing IST treatment and evaluatees opined IST who have not yet received IST treatment.

Taken together, research on non-content-based validity scales suggest that symptomology as well as intellectual ability levels can impact the ability to respond consistently on the MMPI. Though the MMPI-3 is designed to be accessible to individuals with varying abilities and backgrounds, consistent completion of the measure also requires some degree of cognitive skill and resilience (a fourth grade reading level). Graham (2006) pointed out that the MMPI can be an especially challenging and/or tedious task for individuals with psychological illnesses. The measure places significant cognitive demands on test-takers; low IQ or psychological illnesses can negatively impact an individual's cognitive capacity. Depletion of cognitive resources from active symptomology might partially explain why some defendants are unable to aid in their own defenses appropriately. Ultimately, the non-content-based validity scales may be effective tools in differentiating various categories of individuals within forensic populations.

Influences on Inconsistent Responding. Although intervention is unlikely to influence levels of IQ, medications and psychological treatment have demonstrated effectiveness in competency restoration through reductions in and management of psychiatric symptoms (Herbel & Stelmach, 2007). Nicholson and Kugler (1991) completed a semi-meta-analytic review of studies comparing IST and non-IST defendants. They found that the proliferation of symptoms of serious mental illness was negatively related to competency status. Gu et al. (2017) examined MMPI-2-RF validity scale scores amongst CST forensic inpatients undergoing restorative treatment. They report that the presence of genuine psychiatric symptoms can contribute to invalid profiles on the MMPI-2-RF as measured by non-content-based validity scales.

Additionally, Graham (2006) reported research suggesting that genuinely incompetent individuals are frequently omitted from MMPI research, because many are unable to provide consistent and valid profiles, such that clinical scales and content-based validity scales could be interpreted. These studies suggest that patients with more pronounced symptoms of psychosis might experience greater impairment on cognitively stimulating tasks, thus impacting their ability to respond consistently on the MMPI-3.

Available research suggests that medication and other treatments are successful in reducing symptoms that often play a role in incompetency (i.e., disorganization) as treatment progresses (see Herbel & Stelmach, 2007; Mendoza, 2005), thereby improving individuals' ability to engage cognitive resources more effectively and consistently. This indicates that ongoing interventions for IST may reduce the likelihood that an individual will provide an inconsistent MMPI-3 profile.

Overreporting (F, Fb, Fp, Fs, FBS, RBS).

F and Fb Scales. The F and Fb scales were originally utilized to distinguish between exaggerated and genuine psychopathology (Dahlstrom et al., 1975). Indeed, they consist of items infrequently endorsed by a normative population. The F-r scale has demonstrated a sensitivity of .89 and a specificity of .88–.91 in identifying feigned psychiatric symptoms when compared to the Structured Interview of Reported Symptoms (SIRS; Rogers et al., 1992; Sellbom et al., 2010). Despite the utility of the F scale in identifying overreporting, the scale has been less useful in inpatient settings. Research has illuminated that psychiatric inpatients with genuine psychopathology also frequently provide elevations in F (Arbisi & Ben-Porath, 1995). Problems using the F scale within an inpatient sample are unsurprising, given that the scale items were based on infrequent responding amongst non-clinical samples only (Glassmire et al., 2017).

Since clinical samples are likely to present with more genuine symptoms than non-clinical samples, it is also likely that they would endorse items considered for inclusion in the F scale.

In forensic contexts, the F and Fb scales have performed similarly. In a forensic psychiatric sample, the F-r scale demonstrated high correlations with multiple content scales, suggesting that it is associated with a wide range of psychological problems. Toomey et al. (2009) found that the F and Fb scales were consistent in identifying malingering in a forensic sample when compared with the SIRS (Rogers et al., 1992). Additionally, Glassmire et al. (2017) suggested that the F-r actually functioned better than expected in a forensic population. They found that most items on the F-r scale were endorsed by less than 20% of their sample, suggesting that the improved normative sample contributed to more accurate validity scale performance. Although they demonstrate an ability to detect the presence of overreported psychopathology, they also are likely to generate false-positives within a clinical forensic population. Greene (2011) pointed out that although F and Fb scales are largely analogous in how they were constructed, they do differ in item content. Specifically, items included in the Fb scale address content more severe psychopathological symptoms more infrequently endorsed than the items within the F scale.

In civil forensic settings, the F and Fb scales could have greater utility, depending on the specific nature of the setting. For example, were the MMPI or its revised versions used in a civil forensic setting where psychopathology was not a concern, such as custody cases, the F and Fb scales would have greater utility. However, consistent with inpatient research, forensic inpatients adjudicated not guilty by reason of insanity (NGRI) still endorse F-r items to a greater extent than would be expected (Glassmire et al., 2017). Ben-Porath and Tellegan (2008) suggested that

elevations on F-r in inpatient settings do not necessarily indicate the presence of exaggerated symptomology.

Despite the apparent confounding nature of genuine psychopathology, the F and Fb scales have been cautiously utilized in clinical and forensic settings. The research outlined suggests that though the F and Fb scales are effective in identifying the presence of overreporting, they can be indirectly elevated by the presence of genuine psychopathology. Graham et al. (1991) found that discriminating between inpatients with genuinely elevated F scores from individuals instructed to intentionally feign psychopathology required T scores upwards of 120. Graham et al. (1991) noted that multiple studies have suggested that T-scores above 100 are likely indicative of feigned psychopathology in community samples. The F and Fb scales are likely to have limited utility to in psychiatric forensic settings due to the preponderance of genuine psychopathology. As such, the F and Fb scales may be less useful in making important distinctions between forensic subgroups. Defendants who are incentivized to overreport psychiatric symptoms as well as defendants providing accurate representations of their illnesses may provide similar profiles regarding F and Fb scales. Subscales normed using an inpatient sample might be important. In addition, the research highlights the importance of the evolution of the normative sample. Scale utilizing dated and more limited norms performed differently than updated scale utilizing more recent norms. That demonstrates the importance of maintaining a normative sample that represents populations accurately.

Fp Scale. In response to criticism regarding the use of F and Fb amongst individuals with known severe psychopathology, several subscales used to measure overreporting, such as the Fp scale, were developed to better assess specific populations. The Fp scale is arguably more applicable to inpatient psychiatric populations, because the development of the scale involved

identifying rarely endorsed items across clinical contexts (Arbisi & Ben-Porath, 1995). Specifically, the Fp scale consists of items endorsed by fewer than 20% of inpatients. Subsequent research comparing F with Fp has noted a reduction of invalid scores found in inpatient samples as a result of the development of the Fp scale (LePage & Mogge, 2001). LePage and Mogge (2001) highlighted that the use of the Fp scale in their inpatient sample allowed for an increase in the number of interpretable profiles, given fewer were deemed invalid, ultimately increasing the interpretive utility of more MMIP-2 profiles. They also highlighted that concurrent measures (e.g., Personality Assessment Inventory [PAI]) administered to the same group resulted in significantly fewer invalid protocols than the MMPI-2 overall, suggesting that the F scale may produce false positive indications of overreporting in certain contexts. The use of the Fp scale contributes to similar (though still higher than the PAI) findings of invalid protocols between the MMPI-2 and other similar measures, suggesting that the F and Fb are not always reflective of actual overreporting in inpatient samples.

Though research has also indicated that the Fp scale is particularly useful in forensic contexts, some researchers have questioned its utility in predicting malingered psychopathology in a forensic sample. Kucharski et al. (2004) found that the Fp scale added no incremental validity to the information attained by the F scale. More recent research also concludes that the Fp-r scale and the F-r scale are somewhat unreliable in forensic decisions regarding malingering. Sanchez et al. (2017) found that the Fp-r had the best discriminative ability in clinical contexts, whereas the F-r scale performed better than other scales in general population samples, in line with previous research. However, they also found that both scales contributed to significant numbers of false positive as well as false negative reports of malingering. They suggested that the Fp score could be used as a piece of information in a forensic evaluation, but that it does not

have the empirical support to detect malingering in and of itself. The authors cite the SIRS as a relevant empirical measure that *is* validated to independently detect malingering in a forensic sample. McCusker et al. (2003) examined the comparability between the SIRS and the MMPI-2 validity scales. They determined that although each test maintained unique advantages, they reflect similar scores related to malingering.

Contrasting evidence, however, suggests that the MMPI validity scales have been effective in detecting malingering in criminal forensic contexts. Toomey et al. (2009) found that the Fp scale was consistent in identifying malingering when compared with the performance of the SIRS amongst forensic inpatients. In addition, Wygant et al. (2007) found that the Fp and the F scales were successful in predicting cognitive response biases in addition to psychopathology in forensic evaluations. Sellbom et al. (2010) examined the performance of the MMPI-2-RF validity scales in forensic evaluations. As mentioned previously, they found that the F-r and Fp-r both were predictive of malingering in a forensic sample, with Fp-r as the only validity scale to provide incremental validity to that of the F-r.

Though there is research to support the use of the Fp scale in forensic decisions related to malingering, it remains unclear whether the Fp is useful in distinguishing certain malingering groups in forensic contexts. Specifically, some research suggests that the Fp scale contributes to inaccurate evaluations whereas other research indicates the scale is useful in detecting psychiatric and cognitive malingering. It will be important for the present research to examine whether different forensic subpopulations perform differently on the Fp scale, such that the scale can accurately distinguish malingering from actual dysfunction.

Fs Scale. Wygant et al. (2004) developed the Fs scale, which includes items infrequently endorsed by patient samples with genuine medical conditions and chronic illnesses. Like the Fp

scale, the Fs scale was normed using a more specific and clinical population, making the scale more applicable to inpatient populations. Although frequently used in inpatient medical settings, there is little research to suggest that the Fs scale has utility in distinguishing feigned psychotic disorders. Given that the Fs is a relatively recent addition to the MMPI family, there is more limited information about how the scale performs in different settings.

In a recent meta-analysis, Ingram and Ternes (2016) identified that the Fs scale is highly context specific and does not perform equally within different populations. For example, although the Fs scale adequately detects malingering in medically related evaluations, it is less effective in forensic evaluations. Available research suggests that despite its unique ability to measure somatic complaints specifically, the Fs scale performs similarly to other overreporting scales in certain contexts. Wygant et al. (2009) found that, like F-r and Fp-r, the Fs scale successfully differentiated malingerers in groups of head injury, medical patient, and injury/disability samples. Schroeder et al. (2012) also found that the Fs scale demonstrates reasonable effect sizes within a sample of neuropsychological participants. Available research suggests the Fs scale was successful at distinguishing between patients with nonepileptic seizures and those with epileptic seizures (Locke et al., 2010). Given the correlations observed between the Fs scale and other overreporting scales, the Fs scale has also been examined in forensic contexts.

Sanchez et al. (2017) concluded that the Fs scale had limited utility in a forensic population. Though they found that the Fs scale had an 82.01% sensitivity to detect malingerers in contrast to a community sample, it was less effective at distinguishing malingerers from a clinical population (70.78%). Conversely, Sellbom et al. (2010) compared Fs scale performance with other overreporting scales against the SIRS rating scale in a forensic context. They found

that though less successful in differentiating malingerers versus non-malingerers in a forensic population than other scales of overreporting, the Fs scale produced reasonable effect sizes, which the authors argue demonstrates the scales utility in the forensic context. In addition, Wygant et al. (2007) found that criminal forensic defendants were likely to mangle in multiple domains of functioning including somatic symptoms, whereas civil forensic litigants are likely to present a more precise picture of dysfunction.

The Fs scale has demonstrated incremental validity to assessments in forensic settings. Though the measure has received more limited empirical scrutiny than other scales, it may provide important information in detecting malingering. Wygant et al. (2007) provided evidence that criminal malingerers are often likely to exaggerate somatic symptoms in addition to psychiatric and cognitive symptoms. However, the field remains somewhat inconsistent regarding the extent to which the Fs scale adds incremental validity to other overreporting scales in forensic contexts. As such, it will be important for research examining overreporting in forensic contexts to examine the Fs scale to determine whether differences on scale performance are reflective of forensic decisions about malingering and psychopathology.

FBS Scale. Similar to the F scale, the FBS scale has been criticized for its propensity for false positive identifications of overreporting. Critics have argued that the FBS captures an inappropriate amount of genuine psychological distress, particularly in inpatient samples (see Bucher et al., 2003). Even research examining recent editions of the FBS (FBS-r) have found that the FBS-r is ineffective in distinguishing malingered symptomology amongst clinical populations (Sanchez et al., 2017). Sanchez et al. (2017) examined the performance of the validity indices amongst clinical psychiatric patients, malingerers, and general population participants. They found that all validity scales were effective in distinguishing between

malingers and the general population. However, only F-r and Fp-r were effective at distinguishing between malingerers and the clinical population. The results suggest that not all measures of overreporting are equally effective in distinguishing malingering in certain clinical contexts. Butcher et al. (2008) also criticized the use of the FBS in clinical settings. They found high false positive rates amongst psychiatric inpatients and argue that the measure more accurately represents genuine maladjustment and somatic symptoms. Sanchez et al. (2017) discussed that although the FBS-r performed less favorably in clinical populations than other overreporting scales, the scale still produced a large effect size in distinguishing malingerers in a general population.

As discussed, the FBS scale has been criticized for its limited scope as well as its limited validity in inpatient psychiatric settings. Indeed, the scale focuses heavily on somatic symptoms often not characteristic of psychiatric illnesses. The scale was originally designed for use in civil forensic evaluations (Wygant et al., 2009). Specifically, the FBS has been effective in detecting malingered emotional distress, somatic symptoms, neurocognitive dysfunction, and suboptimal effort in cognitive tasks in civil forensic settings. However, Wygant et al. (2009) found that the FBS-r has at least some utility in criminal forensic settings as well. They found that in addition to civil settings, the FBS outperformed other validity measures in detecting malingered somatic and cognitive symptoms in criminal forensic settings as well. Further, Sellbom et al. (2010) found that the FBS-r produced incremental validity in detecting malingering in forensic settings when compared to the F-r scale and had a medium effect size when used to discriminate malingered symptoms. The FBS-r performed similarly to the Fs scale within a clinical forensic sample. Further, a meta-analytic review suggests that the FBS scale has strong empirical support to detect overreported somatic and cognitive symptoms in forensic contexts (Nelson et al., 2006). Nelson

et al. (2006) conducted analyses of moderators for the different MMPI-2 validity scales. They found that when effort was known to be low and when patients had traumatic brain injuries, the FBS demonstrated larger effect sizes than other overreporting validity scales. Their findings suggest that the FBS is uniquely valuable in certain forensic contexts to determine the extent of somatic reporting. Additionally, research suggests that the FBS-r is minimally impacted by contextual moderating variables (Ingram & Ternes, 2016). The FBS-r scale demonstrates strong discriminability and has demonstrated stronger consistency than some other scales of overreporting.

In sum, research about the utility of the FBS scale are somewhat mixed. In inpatient clinical contexts, the scale has demonstrated small effect sizes and poor discriminability in detecting malingering, as the scale was not developed to detect overreported psychiatric symptoms. However, when applied to various forensic contexts, the scale has added important incremental validity. It appears that the scale is effective in detecting overreports of cognitive and somatic complaints. Previous research suggests that malingerers in clinical forensic contexts frequently overreport in multiple domains of functioning. Therefore, it is understandable that the FBS scale would perform differently in clinical versus clinical forensic contexts. The FBS is arguably an important scale for forensic evaluations. Research is needed to identify the extent to which the FBS is likely to be reflective of forensic decisions to further examine its utility in that context.

RBS Scale. Gervais et al. (2007) discussed that the RBS scale was originally designed as a measure of neurocognitive symptom exaggeration particularly to be used in personal injury or disability samples. As such, research examining the RBS in psychiatric inpatient samples is sparse. However, Ingram and Ternes (2016) found that the RBS demonstrated large effect sizes

in identifying malingerers in multiple contexts. They note that the RBS had few associated moderating variables, which strengthens the stability of the scale. Further, given that the scale has been shown to effectively identify malingerers in multiple contexts and is the validity scale least susceptible to influence from contextual factors, the RBS is likely an effective scale in the context of neuropsychological evaluations. They argue that other validity scales have demonstrated less ability to identify malingered neurocognitive symptoms, making the RBS an important component of evaluations where malingering is concerned.

There is considerably more research about the application of the RBS in a forensic setting. Wygant et al. (2010) composed one of the first studies to consider the performance of the RBS scale in a forensic psychiatric sample. They addressed concerns that the RBS scale did not add incremental validity to other overreporting scales in forensic samples. Their results suggest that the RBS is correlated with multiple validity scales in forensic samples and may tap a subset of individuals who exaggerate both cognitive and psychiatric symptoms, reflecting a novel presentation of symptom exaggeration. More recently, Grossi et al. (2017) examined the RBS within a sample of IST pretrial defendants. They compared performance of the RBS scale with components of the SIRS-2 designed to measure feigned cognitive impairment; they utilized the Test of Memory Malingering (TOMM) as a criterion. Their results suggest that the RBS was effective in identifying cognitive impairment feigning as measured by the TOMM. They caution, however, that the RBS yields a more stringent cut-off than the TOMM, which may indicate the RBS is ineffective in identifying more moderately feigned cognitive symptoms. In their meta-analysis, Ingram and Ternes (2016) suggested that the RBS scale lends utility to forensic evaluations given that it is less susceptible to contextual moderators and that it is useful in detecting particularly extreme forms of neurocognitive malingering. They further highlighted

that other scales have demonstrated a relative lack of consistency in predicting neurocognitive malingering. As such, the RBS is likely to be an important scale within forensic evaluations.

An examination of the performance of the RBS in various contexts revealed that it performs largely similarly to the FBS-r. Although the RBS and FBS-r yield somewhat smaller effect sizes than some other scales, they demonstrate strong consistency across contextual variables, which increase their stability. In addition, like the FBS-r, as well as the Fs scales, critics have questioned whether the scale adds incremental validity to other scales of overreporting. The prevailing literature indicates that, particularly within a forensic psychiatric context, the RBS uniquely identifies overreported neurocognitive symptoms that would not otherwise be consistently identified by other validity scales. It appears that a comprehensive forensic evaluation necessitates consideration of the RBS scale as evaluators determine the extent to which overreporting impacts MMPI performance.

Overreporting and Evaluatee/Inpatient Status. Sellbom et al. (2010) found that a group of forensic evaluatees who were known to have feigned psychopathology (per validated psychological assessment tools used to detect feigned psychological symptoms) provided higher scores on F-r and Fp-r when administered the MMPI-2-RF. In addition, Glassmire et al. (2017) found that when not incentivized to overreport, post-adjudication forensic inpatients did not display elevations on the F-r and Fp-r scales. This research suggests that individuals within a forensic setting are likely to provide higher scores on F and Fp during a forensic evaluation than in other contexts. These findings are understandable given the difference in external pressures between pre- and post-adjudication. Wygant et al. (2007) found that the F and Fp scales accurately distinguished individuals who feigned or experienced genuine psychopathology in forensic evaluations. They discuss that forensic evaluatees likely experience uniquely high

pressure to feign psychological impairment. Therefore, forensic evaluatees appear more likely to experience external pressures to feign symptoms of psychopathology than forensic inpatients with few external pressures. Thus, they would be likely to display higher scores on F and Fp on the MMPI-3.

Underreporting Scales (L & K). Scales of underreporting have been investigated with various populations. However, few have examined them strictly with forensic inpatient psychiatric samples; most have examined outpatient clinical samples. The examination of inpatient populations in contrast to outpatient populations could provide insight about how inpatients and outpatients perform differently. Often, inpatient status is involuntary and involves limitations to freedoms. As such, an examination of self-favorable bias is an important component of an inpatient evaluation.

Much of the work investigating the performance of underreporting scales involves community or civil forensic populations (Bagby et al., 1997). Detrick and Chibnall (2014) investigated the extent to which specific contexts impact the underreporting validity scales on the MMPI-2-RF. They found elevations in validity scales measuring underreporting (*L & K*) were often elevated during high-stakes employment selection processes. Similarly, Archer et al. (2012) found that in custody proceedings, parents subjected to the MMPI-2 and MMPI-2-RF were likely to display elevations on scales L and K.

Bagby et al. (1997) examined the performance of underreporting scales on psychiatric patients. They compared concurrent performances of psychiatric inpatients that were asked to either respond honestly or to underreport their symptomology. They concluded that psychiatric patients diagnosed with schizophrenia performed similarly to community members on measures of underreporting when asked to simulate faking-good. The researchers discussed that their study

produced elevations in both L and K for patients diagnosed with schizophrenia who were asked to simulate under-reporting. Similar to a meta-analysis examining the performance of L and K on the MMPI-2 (Baer & Miller, 2002), Sellbom and Bagby (2008) found that L and K successfully differentiate between intentional underreporting and honest responding on the MMPI-2-RF. Further, they found that elevations in underreporting were also associated with lower symptomology reported on clinical scales, suggesting that the L and K validity scales are effective in identifying inhibited reporting of symptomology.

In forensic contexts, L and K scales are frequently considered in situations where evaluatees are considered for child custody or conditional release (Sellbom & Bagby, 2008). In those contexts, evaluatees are often motivated to present themselves favorably, such that they can influence a favorable outcome.

Underreporting and Evaluatee/Inpatient Status. As indicated previously, evaluative settings, particularly those that are a part of the judicial process, can influence evaluatees to present themselves as faking good. There are two important components to faking good on the MMPI L scale. Some MMPI test-takers desire to display good adjustment and few symptoms of pathology. In contrast, the L scale also consists of items designed to capture attempts to appear unrealistically honest, moral, or conforming. In civil forensic settings, MMPI respondents often display elevations in L due to the desire to present themselves as low in pathology and as highly virtuous (Melton et al., 2018). Indeed, civil litigation, such as parental custody, hinges on parents' abilities to present themselves as high functioning and highly virtuous. In criminal forensic settings, however, defendants may be only incentivized to present themselves as virtuous, but not necessarily well-adjusted or high functioning. Research has also examined the L scale in clinical samples with varying external pressures.

Sellbom and Bagby (2008) found that the L-r scale on the MMPI-2-RF was successful at distinguishing feigned uncommon virtues between patients diagnosed with schizophrenia who were engaged in a research study with different external pressures. Individuals asked to intentionally underreport provided significantly higher scores on L-r than those provided standard instructions on the MMPI-2-RF. Additionally, that study demonstrated that clinical patients engaged in a research study with no incentive to manipulate their response styles provided average scores on the L-r scale, indicating they generally did not report uncommon virtues. That suggests that criminal forensic psychiatric inpatients are not inherently likely to provide elevations on L without the presence of external motivating factors.

Though there is research indicating that individuals engaged in civil forensic litigation and clinical settings commonly provide elevations on the L scale when they are incentivized to present themselves in a positive light (see Richey & Doninger, 2020), Melton et al. (2018) reports that there is limited research examining the L scale in criminal forensic settings. However, Bagbey et al. (1995) found that the L scale is effective at distinguishing forensic inpatients who completed the MMPI-2 as a part of a low-stakes non-forensic intake assessment from individuals with more overt pressures to present themselves in a positive light. Their results suggest that forensic inpatients with few external pressures will not produce elevations on the L scale, whereas those undergoing a forensic evaluation may have greater incentive to present themselves as overly virtuous. Criminal forensic evaluatees may be unlikely to display clinically meaningful elevations in the L scale per research suggesting criminal forensic evaluatees are more likely to exaggerate the presence of psychopathology rather than to suppress psychopathology. However, the pressure to present themselves as virtuous or as a good people may contribute to higher scores on L than for individuals who do not have any vested interest in the outcome of the

MMPI. There was no research found examining forensic inpatients performance on MMPI instruments who knowingly completed the MMPI as a part of a research study. However, the aforementioned research suggests that, given their investment in the outcome of an evaluation, forensic evaluatees would be likely to display higher scores on the L scale than forensic inpatients involved in a research study, who have no vested interest in the outcome of the MMPI results.

Overall, available research suggests that the validity scales on the MMPI play important roles in measuring different forms of response bias within forensic contexts. Although each scale has apparent limitations and several scales have limited associated research in clinical forensic samples, they each appear to add incremental validity to an overall profile of response style in a forensic evaluation. Further, it appears important to examine difference in response bias between different forensic subpopulations. More research is needed to identify how different situational contexts impact performance on validity scales within forensic settings. Further insight could influence clinical expectations for certain subgroups of a forensic population.

Gaps in the Literature

In forensic contexts, MMPI instruments have been commonly used in evaluations pertaining to competency to stand trial (CST) and criminal responsibility (CR) evaluations. Limited consideration has been given to the unique pressures that can impact response bias for different forensic populations. Whereas MMPI research frequently describes the forensic population used (i.e., pretrial, NGRI adjudicates), few studies have provided comparative data examining how individuals perform differently pretrial or post trial. Despite relatively scant literature, that which is available suggests that different situational factors inherent to pretrial or post trial contexts contribute to discrete response styles that can be measured by MMPI instruments. Wasyliv et al. (1988) found that individuals undergoing an evaluation for criminal

insanity and/or competency to stand trial were more likely to provide MMPI profiles indicative of malingering than individuals already adjudicated NGRI. Differences between the samples as well as contextual differences were likely important factors that contributed to their results. Specifically, the NGRI group had few contextual pressures to provide invalid MMPI profiles. In contrast, individuals still undergoing evaluation were likely motivated to achieve a certain evaluation result, such that it could lead to a favorable judicial outcome. The results found in Wasyliw et al. (1988) suggested that even sub-contexts within forensic psychiatric settings can differentially impact how individuals approach psychological testing. Rogers (2018) discussed external incentives for malingering have a positive relation with rates of malingering. The incentive among NGRI litigants to malingering appears to reduce following adjudication. Indeed, research has found few differences in MMPI-2 validity scale scores between NGRI adjudicates and civil psychiatric patients (see Moskowitz et al., 1999). Overall, it appears important for research to distinguish response styles that are likely present in pretrial and post trial defendants.

Another limitation of previous literature is that many studies examining pre-trial and post adjudication NGRI evaluatees do not discriminate pre-trial defendants based on their evaluation outcomes. Rather, they consider NGRI/IST evaluatees as a whole group. As indicated by the research discussed, individuals involved in different components of the legal process (CST, CR) likely have different external pressures and psychiatric symptomology that impact their response bias.

In addition, forensic psychiatric samples have consisted of both individuals who are and are not ultimately opined competent to aid in their own defenses. As previously discussed, Ben-Porath (2012b) suggested that cognitive or intellectual deficits may contribute to elevations on certain validity scales on MMPI instruments. Individuals opined IST often experience

psychological limitations that prevent them from aiding in their own legal defenses. Prior to the implementation of psychiatric intervention, psychotic disorders, which are often subsequently used as a basis for NGRI pleas, can impact competency. Intellectual functioning can also affect CST (Nicholson et al., 1988). As such, individuals who are opined IST are likely to have difficulty completing tasks such as MMPI instruments effectively. Research on MMPI instruments suggests that individuals with impairments such as those seen in IST defendants are likely to demonstrate variable or fixed response styles. In contrast, patients adjudicated NGRI and undergoing symptom management have been previously deemed competent. Additionally, NGRI inpatients likely have different motivational influences impacting their approach to assessments from patients opined IST. The external pressures and psychiatric symptomology characteristic in IST defendants appears to uniquely impact their approach to psychological assessments. An examination of observed situational differences between IST defendants and individuals adjudicated NGRI might illuminate a more nuanced picture of how forensic samples are likely to approach MMPI instruments.

More recently, Sellbom (2017) argued that groups used to provide normative expectations for forensic groups are overly heterogeneous and provide an inaccurate representation of how different subgroups are likely to perform on the MMPI-2-RF. He found that, on average, when examined separately, CST and CR evaluatees produced significantly different profiles on validity scales. The results provide evidence for examiners to compare their findings with different subpopulations in forensic settings. The results from Sellbom (2017) allow for examiners to determine the extent to which MMPI-2-RF results are typical or atypical. Importantly, Sellbom's (2017) work reflects an examination of mean scores of individuals currently under evaluation for either CST or CR. However, his study does not directly examine

differences between individuals opined IST, competent, criminally responsible, or NGRI.

Additionally, the research does not examine the impact of situational influences on MMPI-2-RF performance, such as impending litigation or treatment intervention. It appears that little research has been devoted to such a topic, highlighting a need for additional exploration.

There is relatively limited research comparing differences in response styles between different forensic subgroups. The current research project will provide information about how different pressures can impact test validity on the MMPI-3. In addition, the current research will inform about possible threats to validity on the MMPI-3 for individuals at different stages of the legal process. The information gained will help evaluators and practitioners better understand the different influences on response style in forensic populations.

Hypotheses

Terminology

The following abbreviations will be used to describe different legal terminology:

- CR = Criminal Responsibility: An evaluation in Michigan to determine if, as a result of mental illness as defined in section 400 of the mental health code, 1974 PA 258, MCL 330.1400, or as a result of having an intellectual disability as defined in section 100b of the mental health code, 1974 PA 258, MCL 330.1100b, that person lacks substantial capacity either to appreciate the nature and quality or the wrongfulness of his or her conduct or to conform his or her conduct to the requirements of the law
- CST = Competency to Stand Trial (An evaluation in Michigan to determine if, as a result of a mental condition, a defendant is capable of understanding the nature and object of the proceedings against him or of assisting his defense in a rational manner)
- IST = Incompetent to Stand Trial
- NGRI = Not Guilty by Reason of Insanity

The following terms will be used to define the different participant groups considered in this study:

- Evaluatees = Outpatient defendants undergoing an evaluation for either CST or CR
- CR Evaluatees = Outpatient CR evaluatees opined criminally responsible
- CST Evaluatees = Outpatient evaluatees opined competent to stand trial
- IST Evaluatees = Outpatient evaluatees opined IST
- NGRI Evaluatees = Outpatient CR evaluatees opined NGRI
- Inpatients = Patients located at the CFP undergoing treatment for IST or NGRI
- IST Inpatients = Inpatients undergoing competency restoration treatment

- NGRI Inpatients = Inpatients adjudicated NGRI

The following abbreviations reflect MMPI-3 Validity Scale variables utilized in this study:

- CNS – Number of items omitted
- VRIN – Random responding
- TRIN – True/False response bias (higher = “True” response bias)
- CRIN – Composite of non-content-based response style (i.e., VRIN and TRIN)
- F – Infrequently endorsed symptoms of psychopathology (community sample)
- Fb – Infrequently endorsed symptoms of psychopathology (2nd half of test)
- Fp – Infrequently endorsed symptoms of psychopathology (psychiatric patient sample)
- Fs – Infrequently endorsed somatic symptoms
- FBS – Infrequently endorsed cognitive or somatic impairments/symptoms
- RBS – Infrequently endorsed cognitive impairments
- L – Infrequently endorsed virtues
- K – Denial of psychopathology

Hypothesis 1

IST evaluees and IST inpatients will provide higher scores on VRIN, TRIN (greater distance from 50T), and CRIN than NGRI evaluees or NGRI inpatients. The research noted above suggests that individuals who are IST are likely to have impairments that could impact their ability to respond consistently, whereas individuals opined or found NGRI will have fewer impairments, due to an increased stability of intrusive symptoms and a lower likelihood of a low IQ.

Hypothesis 2

IST evaluatees will provide higher scores on VRIN, TRIN (greater distance from 50T), and CRIN than IST inpatients who have normal speech content (SC). The aforementioned research indicates that IST defendants become less disorganized as they receive treatment, which may influence their ability to provide consistent responses on the MMPI-3.

Given that some IST inpatients may not display significant improvements in functioning due to persistent mental illness or cognitive dysfunction, a variable to measure mental status will be included. Specifically, IST inpatients will be separated into those who displayed normal speech content, and those who displayed other than normal speech content in the psychiatric note most recently prior to the MMPI-3 administration date. As such, the analyses will examine differences on VRIN, TRIN, and CRIN between IST inpatients with normal speech content (SC), IST inpatients with non-normal speech content, and evaluatees opined IST.

Hypothesis 3

Evaluees will provide higher scores on the F and Fp scale than forensic inpatients. The aforementioned research suggests that individuals undergoing a forensic evaluation are likely to experience external pressure to achieve a certain evaluation outcome, whereas individuals with no external pressures are less likely to exaggerate symptoms of a mental illness. Specifically, the research suggests that criminal forensic evaluatees often exaggerate the presence of psychopathology.

Hypothesis 4

Evaluees will provide higher scores on the L-Scale than inpatients. Research suggests that psychiatric patients will provide no elevation on the L-scale when they are provided normal MMPI instructions absent external incentive to “fake good.” Research suggests that civil forensic

evaluatees are often motivated to present themselves in a positive light, and are thus, likely to provide elevations on the L-scale. They are influenced to not only present themselves as free of psychopathology, but also highly virtuous. It was predicted that though criminal forensic evaluatees would be incentivized to overreport psychopathology, they would still be incentivized to display themselves as highly virtuous to a greater extent than the inpatient sample. Forensic inpatients taking part in a research study are likely to have little motivation to present themselves in an overly virtuous light, in contrast to individuals undergoing a forensic evaluation. To date, there are no studies examining this difference. It is anticipated that though neither group will produce clinically significant elevations, they will display a statistical difference on the L scale.

The results of the present study will provide information about how the MMPI-3 validity scales discriminate individuals in different forensic contexts. They will serve as a foundational basis for the development of normative expectations regarding how certain individuals are likely to approach psychological evaluations under specific conditions.

Method

Data, Data Storage, and Variables

The present involved an archival analysis of existing records and data at the Center for Forensic Psychiatry (CFP). Experimental procedures were not conducted as direct part of this study. The research data utilized were collected as a part of a separate inpatient validation study of the MMPI-3, in which the author of the current study was directly involved. Specifically, the author of the present study was involved in the design, organization, administration, data collection, and data analysis for the study. The data was obtained following a data use agreement between this author and the State of Michigan for purposes of the present research.

This author had no access to identifiable data. The data were compiled and de-identified by CFP staff members. The data set was saved on an encrypted flash-drive provided by the CFP to be used by the authors of this study until the completion of data analysis. Following data analysis, the flash-drive was returned to the CFP to be erased. During data analysis, only members of the research team had access to the data.

The procedure for data collection is denoted below to provide clarity to the reader regarding the nature of the dataset as well as the populations examined. It should be noted that data collection was conducted under the guise of an IRB-approved research study at the CFP.

Population and Setting

The present study was conducted at the CFP, an eight-unit maximum-security inpatient forensic psychiatric facility. The CFP is located in Saline, Michigan, and provides evaluation and treatment services to defendants with mental illnesses or who are found IST and NGRI adjudicates from across the state of Michigan. Non-resident defendants are frequently transported from institutions of incarceration for various types of evaluations at the CFP. In

addition, defendants who have been granted bond obtain their own transportation to the CFP for evaluation. An evaluation unit (EU) is designated at the CFP for conducting evaluations regarding criminal responsibility, CST, and other psycho-legal questions. Evaluators at the CFP are tasked to provide opinions with regard to various psycho-legal questions. The results of their evaluations are considered by judges and/or juries who ultimately determine whether a defendant is criminally responsible, NGRI, or IST.

The CFP inpatient population consists of male (7 units) and female (1 unit) forensic inpatients designated IST, IST probate, or adjudicated NGRI. The inpatient population ranges in age from 18 to upwards of 70 years old. Additionally, the CFP evaluates and treats individuals with various demographic backgrounds. Patients adjudicated NGRI are referred to the CFP on a 60-day diagnostic order to determine the extent of their mental illness as well as their need for treatment. At the end of that order, an evaluator makes recommendations about whether the individual is a danger to themselves or others, and whether they should be treated in a hospital, as an outpatient, or discharged. Most patients undergo treatment at the CFP for longer than the 60-day minimum before they are referred for transfer to either a less restrictive institution or community placement. Their treatment progress is monitored by a treatment team consisting of psychiatrists, psychologists, social workers, recreational therapists, occupational therapists, and medical staff members. At the CFP, patients are provided psychological treatment services that occur both individually and within different treatment groups. Patients receive medications related to their mental illnesses as well as education regarding the importance of medication adherence.

Patients found IST are referred for competency restoration treatment for a total period of 15 months or 1/3 of the maximum sentence the defendant could receive if convicted of the

charges against them, whichever is lesser. In the event IST patients are not restored to competency within the statutorily defined timeframe, their charges are dropped. Then it is determined whether they will be civilly committed to the state hospital for continued treatment. Patients who are not restored to competency within the established timeframe and who are civilly committed are then considered “IST probate.” It is important to highlight that although IST inpatients and IST probated inpatients share similarities regarding their impairments, they also bear unique qualities. There is limited research specifically examining differences between restorable and unrestorable IST patients. However, in consideration of their differential responsiveness to treatment, it makes sense to note that IST probate patients have impairing psychiatric or cognitive problems that are resistant to psychiatric and psychological forms of treatment. The current sample including IST inpatients is comprised of both IST and IST probate inpatients. Despite the inherent differences of these two groups, they were incorporated into one single group of IST inpatients for the purposes of this study. In doing so, the authors were able to secure a more desirable sample size and provide foundational data regarding the research questions. It is anticipated and recommended that future research consider examining differences between these subgroups to further illuminate unique threats to assessment validity.

In addition to inpatient NGRI adjudicates and IST defendants, individuals included for the present study are composed partially of outpatient defendants evaluated pretrial and pre-IST designation on the EU. The group is likely to be composed of individuals with varied and feigned psychopathology given they have not previously been evaluated for their alleged offenses.

Informed Consent

Consent to participate was assessed for all inpatients through a semi-structured interview procedure. Research assistants assessed capacity to consent to ensure participants understood the risks and benefits of their participation. Participants evaluated to have sufficient capacity to consent read and signed a consent form as well as an authentication to disclose protected health information for provide permission to get information from their medical records. In the event that patients who were not legally able to consent were still interested in participating, consent was also requested from a legal guardian for participation and for the release of participant medical records held by the CFP.

Consent for participation was not collected from EU participants. Evaluators may request the use of a MMPI-2-RF-EX or a MMPI-3 as a part of standard evaluation procedures. EU data can be requested from the State of Michigan to be analyzed archivally for research purposes. State and federal guidelines assert that information obtained through forensic evaluations for research purposes should be protected to the degree it does not interfere with judicial proceedings. Information collected for the present study was deidentified and presented quantitatively.

Procedure

Inpatient participants underwent a structured research study protocol. Patients were first solicited for participation on their individual units. Researchers read a script that outlined the purpose of the study, procedures, risks, benefits, and the protections for patients. Importantly, patients were informed that their performance would have no bearing on their treatment or their legal standing. They were informed that their individual data would not be shared with members of their treatment or legal teams. Their participation was voluntary and their decision to

participate did not impact the course of their treatment at the CFP. During the initial introduction, patients were invited to voluntarily sign up to participate with the knowledge that they could revoke their interest in participating at any point.

Following initial solicitation, researchers evaluated capacity to consent. Participants who ultimately consented were then evaluated for reading ability. Researchers evaluated each prospective participant individually using the Word Reading subtest of the WRAT5, a psychometrically validated measure of reading ability. Researchers were trained to competency to administer and score specific subtests of the WRAT5 at the CFP prior to the commencement of the study. Data from individuals who had a reading ability below a fifth-grade level as determined by the WRAT5 were excluded from the MMPI-3 study, though they were still provided the incentive.

Participants who consented and had at least a fifth-grade reading level were then subject to a records review similar to that implemented for outpatient defendants. A separate Inpatient Forensic Record Review Form was completed for each inpatient participant. Information pertaining to patient legal status as well as other demographic information was obtained from patient electronic medical records.

Following screening and records reviews, researchers administered the MMPI-2-RF-EX or the MMPI-3 to groups of approximately 10 participants. The MMPI-2-RF-EX was used prior to the release of the MMPI-3. As such, the MMPI-3 was scored for each participant. The MMPI-2-RF-EX/MMPI-3 was administered utilizing standard administration procedures derived from previous editions of the MMPI (see Ben-Porath & Tellegen, 2008).

Inpatient participants were compensated for their participation in the form of a food incentive. The food was administered immediately following the completion of the MMPI-2-RF-

EX or MMPI-3 measure. Prospective participants were informed of the incentive at the time they signed up. Participants who were disqualified from participation during the initial screening process were offered the same food incentive for their willingness to participate.

EU participants were not administered a study-specific participation protocol. Given that forensic examiners at the CFP routinely administer MMPI measures during their evaluations, data was collected from a data pool located at the CFP. As a part of forensic evaluations involving MMPI instruments, examiners first identify that defendants have an adequate reading level. Therefore, explicit measures to evaluate participants' reading levels were not necessary for EU defendants. Demographic information including legal status was obtained from the data pool.

Statistical Analyses

A statistical correction for statistical significance was conducted to account for multiple family-wise comparisons. A Bonferroni adjustment analysis suggests that an alpha of .01 should be used based on the number of comparisons proposed. An analysis using G*Power revealed that to achieve a moderate to strong effect size as well as a power ($1 - \beta$ err prob) of .8, the present study sought to achieve a sample of approximately 102 participants.

For analyses of content-based validity scales, participants who provide scores on CNS (raw score) > 17 or T scores > 79 on CRIN, VRIN, or TRIN were excluded from analyses examining content-based validity scales; elevations on non-content-based validity scales suggests a MMPI-3 profile is not interpretable due to an inconsistent pattern of responding. Group performance on specific validity scales was examined in comparison with other groups within the study. More specifically, Cohen's d values of at least .5 will indicate meaningful differences between group means on validity indices.

The present study employed multiple students' *t*-tests. *T*-tests are used to distinguish differences between two groups' performance on a dependent variable. For the present study, different MMPI-3 validity scale indices served as dependent variables for the different groups analyzed. Tests of between-group comparisons of means commonly have an assumption of homogeneity of variance, which can be difficult to meet when samples are small or of different sizes. The assumption of homogeneity of variance between groups was tested using Levene's test. When the assumption was met, group differences in means were tested with a student's *t*-test, and when it was not met, the Welch's correction to the *t*-test was interpreted.

Hypothesis 1

IST evaluatees and inpatients will provide higher scores on VRIN and CRIN, and a more extreme score on TRIN than NGRI evaluatees and inpatients.

- Independent Samples *t*-test:
 - IV = IST/NGRI Opinion/Status
 - DV= VRIN Score
- Independent Samples *t*-test:
 - IV = IST/NGRI Opinion/Status
 - DV= TRIN Score
- Independent Samples *t*-test:
 - IV = IST/NGRI Opinion/Status
 - DV= CRIN Score

Hypothesis 2

IST Evaluatees will provide higher scores on VRIN, TRIN, and CRIN than IST inpatients who have normal speech content (SC).

- Independent Samples *t*-test:
 - IV = IST Inpatient Normal SC/IST Evaluee
 - DV= VRIN Score
- Independent Samples *t*-test:
 - IV = IST Inpatient Normal SC/IST Evaluee
 - DV= TRIN Score
- Independent Samples *t*-test:
 - IV = IST Inpatient Normal SC/IST Evaluee
 - DV= CRIN Score

Hypothesis 3

Evaluees will provide higher scores on the F and Fp scale than forensic inpatients.

- Independent Samples *t*-test:
 - IV = Inpatient/Outpatient Status
 - DV= F Score
- Independent Samples *t*-test:
 - IV = Inpatient/Outpatient Status
 - DV= Fp Score

Hypothesis 4

Evaluees will provide higher scores on the L-Scale than inpatients.

- Independent Samples *t*-test:
 - IV = Inpatient/Evaluee Status
 - DV= L Scale Score

Results

Descriptive statistics and frequencies were generated and the data were analyzed.

Descriptive statistics and frequencies are denoted in Tables 1 and 2. Table 3 includes means and standard deviations of MMPI-3 scores across the different groups.

Table 1

Demographic Frequencies

		<i>N</i>
CFP Status	Inpatients	114
	Evaluees	41
Gender	Male	129
	Female	19
Race	American Indian/Alaskan	
	Native	1
	Asian	1
	Black/AA	44
	White	99
	Other	2
Oriented	Yes	112
	No	0
	Unknown	36
Speech Content	Normal/Goal Directed	113
	Loose Associations	1
	Bizzare	8
	Tangential	8
	Circumstantial	8
	Disorganized	10
Number of Incident Reports Among Inpatients	0	73
	1	15
	2	5
	3	5
	4	2
	6	4
	13	1
	14	1
18	1	

Table 2*Descriptive Means*

	Mean (SD)	Minimum	Maximum
Age	39 (12.17)	18	68
Education (year)	12.39 (1.97)	8	20

Table 3*Group Means and Standard Deviations*

	IST Participants	NGRI Participants	IST Evaluees	IST Inpatients	IST Inpatients NSC	Inpatients	Evaluees
MMPI-3 Validity Scale	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
VRIN	63.62 (16.68)	61.26 (16.55)	53.17 (10.75)	64.99 (16.91)	59.93 (14.26)	55.92 (17.03)	63.09 (17.03)
TRIN	63.55 (13.58)	64.70 (15.35)	66.08 (16.91)	63.22 (13.56)	59.40 (9.66)	63.50 (14.39)	61.57 (10.23)
CRIN	66.23 (16.52)	62.68 (17.41)	58.98 (14.25)	67.17 (16.70)	61.45 (12.92)	64.08 (17.60)	58.18 (14.59)
F						62.45 (17.77)	80.27 (26.35)
Fp						69.81 (25.78)	74.84 (33.34)
Fs						58.74 (14.54)	68.02 (19.65)
FBS						53.86 (10.50)	67.74 (13.99)
RBS						57.63 (13.40)	72.46 (18.37)
L						56.46 (10.80)	54.02 (7.73)
K						51.82 (10.76)	45.24 (9.98)

Note. IST = Incompetent to Stand Trial; NGRI = Not Guilty by Reason of Insanity; IST Inpatients NSC = Inpatients found IST who have Normal Speech Content

Hypothesis 1

Hypothesis 1 examined whether IST evaluees and inpatients would provide higher scores on VRIN or CRIN, or a more extreme score on TRIN than NGRI evaluees and inpatients. It was hypothesized that evaluees opined IST or inpatients who are undergoing competency restoration ($n = 52$) are likely to have cognitive limitations or psychiatric impairments that impair their ability to attend to and respond consistently on the MMPI-3 to a greater extent than evaluees who were opined NGRI or NGRI inpatients ($n = 60$).

An independent samples t -test was conducted and suggests there was no difference between evaluees and inpatients opined IST or who are undergoing competency restoration ($M = 63.62$, $SD = 16.68$) and evaluees who were opined NGRI or NGRI inpatients ($M = 61.26$, $SD = 16.55$) on VRIN, $t(110) = .75$, $p = .46$, Cohen's $d = .142$. Another independent samples t -test was conducted and suggests there was no difference between evaluees and inpatients opined IST

or who are undergoing competency restoration ($M = 63.55$, $SD = 13.58$) and evaluatees who were opined NGRI or who are NGRI inpatients ($M = 64.70$, $SD = 15.35$) on TRIN, $t(110) = -.42$, $p = .68$, Cohen's $d = -.08$. Another independent samples t -test was conducted and suggests there was no difference between evaluatees and inpatients opined IST or who are undergoing competency restoration ($M = 66.23$, $SD = 16.52$) and evaluatees who were opined NGRI or who are NGRI inpatients ($M = 62.68$, $SD = 17.41$) on CRIN, $F(1,110) = 1.10$, $p = .27$, Cohen's $d = .21$. Taken together, the results of Hypothesis 1 suggest that IST evaluatees and inpatients do not necessarily display more inconsistent responding on the MMPI-3, nor are they more likely to engage in a true or false response bias than NGRI evaluatees and inpatients.

Hypothesis 2

Hypothesis 2 aimed to further describe the response style of evaluatees and inpatients opined or found IST. Specifically, it was hypothesized that IST evaluatees ($n = 6$) would be less stabilized than IST inpatients ($n = 46$) who have undergone competency restoration treatment and who have received psychiatric treatment and, therefore, would display higher scores on VRIN and CRIN, and display more extreme scores on TRIN. Importantly, the nature of the sample collected did not allow for equal comparison groups in the present analysis. Specifically, the IST evaluatee group had only six participants. Regardless, an analysis was conducted as a preliminary estimate of a potential distinction between the groups analyzed. Results will be considered cautiously.

An independent samples t -test was conducted and suggests there was no significant difference between IST evaluatees ($M = 53.17$, $SD = 10.75$) and IST inpatients ($M = 64.99$, $SD = 16.91$) on VRIN, $t(50) = -1.66$, $p = .10$, Cohen's $d = -.72$. Another independent samples t -test was conducted and suggests there was no significant difference between IST evaluatees ($M =$

66.08, $SD = 16.91$) and IST inpatients ($M = 63.22$, $SD = 13.56$) on TRIN, $t(50) = .48$, $p = .63$, Cohen's $d = .21$. Another independent samples t -test was conducted and suggests there was no significant difference between IST evaluatees ($M = 58.98$, $SD = 14.25$) and IST inpatients ($M = 67.17$, $SD = 16.70$) on CRIN, $t(50) = -1.15$, $p = .26$, Cohen's $d = -.49$). Ultimately, the results suggest that IST evaluatees are not necessarily more likely to display more inconsistent responding on the MMPI-3, nor are they more likely to display a true or false response bias than IST inpatients.

As indicated earlier, IST defendants often become less disorganized as they receive treatment, which may influence their ability to provide consistent responses on the MMPI-3. However, that improvement does not necessarily take place immediately upon admission and thus, true differences between IST evaluatees ($n = 6$) and IST inpatients ($n = 46$) who have benefitted from treatment may not have been detected by the previous analyses. As such, speech content was used to exclude inpatients who may still be experiencing high symptomology or who are not responding to restoration treatment. Specifically, inpatient participants with atypical speech content were excluded from this analysis.

An independent samples t -test was conducted and suggests there was no significant difference between IST evaluatees ($M = 53.17$, $SD = 10.76$) and IST inpatients with normal speech content ($M = 59.93$, $SD = 14.26$) on VRIN, $t(28) = -1.08$, $p = .29$, Cohen's $d = -.49$. Another independent samples t -test was conducted and suggests there was no significant difference between IST evaluatees ($M = 66.08$, $SD = 14.75$) and IST inpatients with normal speech content ($M = 59.40$, $SD = 9.66$) on TRIN, $t(28) = 1.36$, $p = .18$, Cohen's $d = .62$. Another independent samples t test was conducted and suggests there was no significant difference between IST evaluatees ($M = 58.98$, $SD = 14.25$) and IST inpatients with normal speech content ($M = 61.45$, SD

= 12.92) on CRIN, $t(28) = -.41, p = .68$, Cohen's $d = -.19$. Overall, the results suggest that IST evaluatees are not significantly more likely to engage in inconsistent responding or display a true/false response bias than IST inpatients.

Given that there were few evaluatees within the sample who were opined incompetent, thus impacting the validity of the analysis examining differences between IST evaluatees and IST inpatients, supplemental analyses were conducted between all evaluatees ($n = 41$) and inpatients ($n = 114$) to provide insight into group differences on MMPI-3 non-content-based validity scales. It was further hypothesized that evaluatees largely consisting of CST evaluatees opined competent would demonstrate less impairment than inpatients, and therefore, provide lower scores on VRIN and CRIN, and less extreme scores on TRIN. Several independent samples t tests were conducted.

An independent samples t -test was conducted and suggests there was not a significant difference between evaluatees ($M = 55.92, SD = 17.03$) and inpatients ($M = 63.09, SD = 17.03$) on VRIN, $t(146) = 2.35, p = .02$, Cohen's $d = .43$. Specifically, inpatients did not provide significantly different scores on VRIN than evaluatees.

Another independent samples t -test was conducted and suggests there was no significant difference between inpatients ($M = 63.50, SD = 14.39$) and evaluatees ($M = 61.57, SD = 10.23$) on TRIN, $t(146) = .79, p = .43$, Cohen's $d = .14$.

Another independent samples t -test was conducted and suggests there was not a significant difference between inpatients ($M = 64.08, SD = 17.60$) and evaluatees ($M = 58.18, SD = 14.59$) on CRIN, $t(145) = 4.51, p = .04$, Cohen's $d = .39$. Specifically, inpatients did not provide significantly different scores on CRIN than evaluatees.

Hypotheses 3 & 4

Hypotheses 3 and 4 aimed to further examine differences in response style between inpatients and evaluatees. The aforementioned research suggests that evaluatees undergoing a forensic evaluation may be influenced by response bias differently than forensic inpatients with no external motivating factors. Prior to data analysis, participants who achieved an invalid MMPI-3 protocol due to elevations on CNS, VRIN, or CRIN, or extreme scores on TRIN were excluded from the forthcoming analyses. It was hypothesized (Hypothesis 3) that evaluatees would be incentivized to over report symptoms of psychopathology infrequently reported by the community and psychiatric sample and therefore would provide higher scores on the F and Fp scales. Levene's test for equality of variances was found to be violated for the present analysis concerning the F scale, $F(1,108) = 13.02, p < .001$. Given this violated assumption, a Welch's *t*-statistic not assuming homogeneity of variance was computed. The results suggest that evaluatees achieved significantly higher scores than inpatients on the F scale. Levene's test for equality of variances was not significant for the Fp scale. The results suggest no significant difference between evaluatees and inpatients on the Fp scale. See Table 4.

Likewise, it was predicted (Hypothesis 4) that individuals undergoing a forensic evaluation (evaluatees) would be incentivized to present themselves as overly virtuous to a greater extent than psychiatric inpatients with no external motivating factors. Specifically, it was hypothesized that evaluatees would provide significantly higher scores on L than inpatients. Levene's test for equality of variances was found to be violated for the present analysis concerning the L scale, $F(1,108) = 4.95, p = .03$. Given this violated assumption, a Welch's *t*-statistic not assuming homogeneity of variance was computed. The results suggest there was no significant difference between evaluatees and inpatients on the L scale. See Table 4.

An added exploratory component of the present research involved examining differences between evaluatees and inpatients on the other validity scales as well. As such, a series of independent samples *t*-tests was conducted to determine differences in response style between the two groups. Levene's test for equality of variances was found to be violated for the present analysis concerning the Fs scale, $F(1,108) = 5.28, p = .024$; FBS scale, $F(1,108) = 4.07, p = .046$; and RBS scale, $F(1,108) = 6.87, p = .01$. Given this violated assumption, Welch's *t*-statistics not assuming homogeneity of variance were computed and reported for Fs, FBS, and RBS. In addition, for F, Fs, FBS, and RBS, comparison groups with smaller sample sizes had greater variance than the groups with larger sample sizes. Interestingly, significant differences between groups were found on FBS, RBS, and K scales. See Table 4.

Table 4

Independent Samples t-tests between Inpatients and Evaluatees

MMPI-3 Validity Scale	Inpatient		Evaluatee		<i>t</i>	df	<i>p</i>	Cohen's <i>d</i>
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)				
F	62.45 (17.77)	80.27 (26.35)	-3.56*	44.98	<.001	-.86		
Fp	69.81 (25.78)	74.84 (33.34)	-.86	108	.39	-.18		
Fs	58.74 (14.54)	68.02 (19.65)	-2.44*	47.68	.02	-.57		
FBS	53.86 (10.50)	67.74 (13.99)	-5.12*	48.14	<.001	-1.19		
RBS	57.63 (13.40)	72.46 (18.37)	-4.19*	47.22	<.001	-.99		
L	56.46 (10.80)	54.02 (7.73)	1.34*	83.4	.19	.24		
K	51.82 (10.76)	45.24 (9.98)	3.00	108	.003	.63		

Note. Significance at $p < .05$; * = Welch's *t*-test used

Potential explanations for the significant differences will be discussed in the discussion section. In considering the aforementioned results, although mean differences between groups may be reliable, the standard errors are large, which reduces the power of significance testing. Given the limited sample size and limited statistical power, large effect sizes are required to detect reliable significant differences. Where an effect is found to be significant, caution will be exercised when interpreting because the point estimate may not generalize well to the population. As such, there is a potential threat to external validity.

Discussion

The current study was implemented to gain a more nuanced understanding of how different forensic populations are susceptible to display response biases. Prior research suggests that although individuals undergoing a forensic evaluation (forensic evaluatees) and forensic inpatients likely share clinical and behavioral qualities (e.g., diagnoses, criminal history), they may experience unique pressures that differentially influence how they approach psychological testing (Rogers & Payne, 2006). Given that MMPI instruments have been commonly used to aid in forensic evaluation and testing, it is important to understand how different contexts impact response bias on individuals within the forensic system. As such, the current study examined forensic evaluatees and inpatients at different points of the legal process to better understand these differences. In addition, the recent advent of the MMPI-3 necessitates research to understand how different contexts impact participant performance on the measure. The results suggest that forensic evaluatees may overreport psychological impairment to a higher degree than forensic inpatients, whereas forensic inpatients may overreport successful psychological adjustment to a higher degree than forensic evaluatees. However, overall, the current study found that there were largely no significant differences in non-content-based responding between the examined forensic groups on the MMPI-3. This section will examine the implications of the findings as well as future directions for further study.

Hypothesis 1 sought to explore whether the unique impairments reflected in IST evaluatees and inpatients would impact their ability/tendency to respond consistently on the MMPI-3 in contrast to NGRI evaluatees and inpatients. It was hypothesized that IST evaluatees and IST inpatients likely experience symptoms of severe mental illnesses and/or cognitive deficits that impair their ability to respond consistently on the MMPI-3. In contrast, by definition, NGRI

evaluatees and inpatients have already displayed sufficient thought organization and cognitive abilities such that they have been deemed competent to proceed in their cases (Melton et al., 2018). It was hypothesized that IST evaluatees and IST inpatients would, therefore, provide higher scores on VRIN and CRIN and more extreme scores on TRIN. However, the results instead suggest that IST evaluatees and inpatients do not necessarily display more inconsistent responding on the MMPI-3, nor are they more likely to engage in a true or false response bias than NGRI evaluatees and NGRI inpatients. The results further indicate that neither group displayed clinically significant inconsistency. There are several potential reasons for this finding.

Importantly, an examination of the participants included in the groups utilized for the analysis indicated that not all groups were represented equally. Although each comparison group consisted of both inpatients and evaluatees, the actual number of evaluatees opined NGRI ($n = 7$) or IST ($n = 6$) was very low. As such, the sample analyzed for hypothesis 1 was primarily comprised of inpatients found IST or NGRI. This finding is not surprising given that evaluations for CST and CR frequently yield negative forensic opinions (Melton et al., 2018). Compiling a sizeable sample of NGRI and IST evaluatees can be a difficult task. As such, the results of this hypothesis are likely more reflective of differences between inpatient groups, overall, than groups of IST/NGRI inpatients and IST/NGRI evaluatees together. It is possible that the anticipated effect would be driven largely by groups of IST and NGRI evaluatees given their different presentations and levels of psychiatric stability at that point of the legal process. Future research could examine whether there are differences in non-content-based responding on the MMPI-3 between different groups of evaluatees only, and using a larger sample size.

Given that the sample for Hypothesis 1 consisted primarily of inpatients, it is important to consider why inpatients of different commitment codes would be likely to display no differences

in non-content-based response bias. In consideration of the nature of the administration of the MMPI-3 to the inpatient population in the current study, there was an equal external incentive across the different groups. Inpatients were incentivized to participate, but there was no incentive to engage in any type of intentional response bias. It is interesting to consider whether the same results would be found if the inpatient sample was administered the MMPI-3 as a part of a psycholegal assessment or evaluation. It is conceivable that with motivating factors associated with a psychological assessment that has implications for treatment, legal standing, and/or discharge, NGRI inpatients may demonstrate greater consistency and lower true/false response bias than IST inpatients as predicted. Motivation (or lack thereof) may have played an important role in the null findings observed in Hypothesis 1.

The results of Hypothesis 1 also indicate that IST inpatients may not necessarily experience different degrees of thought disorganization and/or cognitive impairment than NGRI inpatients. As noted above, schizophrenia has been most strongly correlated with findings of IST when compared to other psychological impairments in an inpatient setting (Viljoen et al., 2002). As such, active psychosis frequently plays a role in findings of IST for many IST inpatients. Likewise, research has also indicated that most NGRI adjudicates carry a primary psychotic disorder diagnosis (Melton et al., 2018). It may be that these inpatient groups are not as different as originally hypothesized. Given that the current sample of inpatient participants were members of the same treatment milieu and likely experienced similar impairments, it makes sense that they would display a similar ability to respond consistently on the MMPI-3. A review of the literature indicates that there are few available studies that directly compare functioning between NGRI adjudicates and IST inpatient defendants. Given the aforementioned information, however, the results of Hypothesis 1 may serve as evidence that inpatient IST defendants and

NGRI adjudicates display similar rates of psychiatric stability and ability to engage in cognitively demanding tasks on average. It may be important for future studies to investigate similarities and differences in cognitive abilities between NGRI and IST inpatients more directly using standardized cognitive assessment measures. That could offer evidence in support of the current findings suggesting that they do not have inherent differences in their abilities to respond consistently on the MMPI-3.

Interestingly, an examination of the mean scores indicates that overall, the sample did not display clinically meaningful elevations on VRIN, TRIN, and CRIN (below 70T), suggesting that the sample as a whole displayed consistency in scores and no true/false-response bias (for normative interpretations, see Ben-Porath & Tellegan, 2020). It appears that when not externally motivated (by a legal outcome etc.), inpatients are likely to display relative consistency on the MMPI-3. It is notable, however, that the sample produced mean scores near the upper cut-score of *normal-no concerns* on non-content-based validity scales. An examination of the technical manual for the MMIP-3 indicates that clinical normative samples produced modestly higher scores on non-content-based response bias scales than the college normative sample. Although there is currently no available comparison data for forensic inpatients on the MMPI-3, it is conceivable that normative scores for non-content validity scales are slightly higher in an inpatient setting. Indeed, forensic inpatients often experience multiple factors that can negatively impact response consistency (e.g., psychopathology, low cognitive functioning, motivation, legal circumstances; Rogers, 2018). Establishing normative differences between forensic samples could be an important line of future research.

Hypothesis 2 aimed to further identify whether differences in psychiatric treatment and/or different external motivating factors would impact scores on non-content-based validity scales

on the MMPI-3. Specifically, this hypothesis examined whether IST inpatients and IST evaluatees would display differences on non-content-based validity scales. It was hypothesized that when provided intensive inpatient psychiatric treatment, IST inpatients would have a stronger ability to respond consistently and without fixed responding than IST evaluatees who have yet to undergo psychiatric treatment at the hospital. In addition, IST evaluatees are likely to experience greater pressure to engage in concerted response bias to increase the likelihood they are opined IST. However, the results indicate that IST evaluatees did not display inconsistent or fixed responding to a higher degree than IST inpatients. Additionally, neither group displayed clinically significant inconsistency. That suggests that IST evaluatees do not necessarily have greater impairment due to their mental illness or cognitive ability that impacts their ability to respond consistently. In addition, they also may not experience unique external pressures to display inconsistency as originally hypothesized. In considering the hypotheses and available literature, it is curious that the hypothesized differences were not found. There are several possible explanations that could support the possibility of a Type II Error. Specifically, the small n of the IST evaluatee group may not be representative of the greater IST population. Although the study yielded acceptable effect sizes, there were very few IST evaluatees, in particular. As such, it is difficult to conclude whether the sample is representative of the population. It is possible that with a higher sample size, the results could have reflected the hypotheses.

Alternatively, the results may also be a reflection of the unique motivational factors in the IST inpatient sample. As noted in the results section, IST inpatients produced mildly inflated scores on non-content-based validity scales, though not clinically significant. It is possible that since there was no overt incentive for inpatients to put forth their best effort during the test administration, they displayed somewhat inflated scores on non-content-based validity scales.

Were the participants administered the MMPI-3 as a part of actual psychological evaluations, they may have demonstrated significantly lower scores on those scales. Given the possibility of a Type II Error, additional analyses were conducted.

It is also possible that the lack of differences in consistency found between IST evaluatees and IST inpatients may have been due to the fact that some IST inpatients do not make significant progress toward psychiatric stability and thought organization even with psychiatric treatment. As such, a subsequent analysis omitted inpatients who were described as having non-normal speech content, a frequent metric to help determine whether someone has made progress toward competency (Melton et al., 2018). Despite the modification to the inpatient sample, the results continued to suggest that IST inpatients and IST evaluatees do not display differences on non-content-based validity scales on the MMPI-3. The results continue to indicate that psychological or motivational factors do not influence consistency in responding on the MMPI-3 differently between IST evaluatees and IST inpatients.

As noted above, there is a significant disparity in sample size between the IST evaluatee and IST inpatient groups. Although analyses including IST evaluatees have produced meaningful effect sizes, the low number of participants in the IST evaluatee group suggests the sample may not be an accurate representation of the greater population. As such, the results offered by hypothesis two may not provide useful information about the groups analyzed. Further analyses were conducted to examine differences between evaluatees as a whole and inpatients as a whole on non-content-based validity scales, as it was predicted that this information could illuminate motivational factors unique to each setting without the confound of low sample size. Although not statistically significantly different due to an adjustment for multiple comparisons, the results suggest a trending difference in response bias consistency between evaluatees and inpatients. The

results indicate that evaluatees *may* be more likely to respond consistently on the MMPI-3 than inpatients with no external incentive. Evaluatees in the current sample were administered the MMPI-3 as a part of a real forensic evaluation, whereas inpatients completed the measure knowingly as a part of a research study. As such, evaluatees may have experienced greater motivation to respond consistently than inpatients for whom their MMPI-3 results had no bearing on their future. If this finding were to be replicated and consistently found between the two groups, it would emphasize the influence of external motivation on the validity of MMPI-3 profiles. Further, it would be interesting for future research to utilize an inpatient sample that completes the MMPI-3 as a part of an actual psychological assessment or evaluation.

In consideration of other factors that could contribute to this trend, it is notable that a majority of the evaluatee group was ultimately opined competent to stand trial and/or criminally responsible, in contrast to inpatients. This is understandable given that multiple studies have found that of those referred for a CST evaluation, only between 20 and 30% are ultimately opined IST (Melton et al., 2018). As such, it is possible that evaluatees in the current sample did not experience the same level of psychological impairment as inpatients. Given the opinions rendered in their evaluations, it is reasonable to conclude that the evaluatees in the current sample had fewer psychiatric symptoms that interfered with their ability to respond consistently on the MMPI-3 than inpatients found IST or NGRI. Therefore, it makes sense that they would display less inconsistency or fixed responding than inpatients with potentially more severe impairments.

An alternative explanation for why the hypothesized differences were not found involves test administration decisions within the evaluatee group. Despite the benefit of psychological testing in a forensic evaluation, testing is not always warranted and can, at times, be contraindicated for a multitude of reasons (Melton et al., 2018). For example, testing may be

contraindicated in evaluations where initial clinical interviews suggest an evaluatee will be unable or unwilling to participate such as to produce meaningful test data. In addition, assessments such as the MMPI-3 require valuable resources (e.g., time, expertise, funds). Therefore, it is possible that the current sample of evaluatees does not include data from individuals with overtly obvious severe psychiatric, cognitive, or personality impairments. Rather, the sample reflects evaluatees who displayed higher functioning and cooperation during forensic interviews. Future studies could consider employing a variable that evaluates level of functioning and cooperation as a covariate to determine predictive trends at different levels of functioning and cooperation.

Hypothesis 3 sought to determine whether evaluatees would be likely to overreport psychological and psychiatric dysfunction to a greater extent than inpatients. Research suggests that individuals undergoing a forensic evaluation may be motivated to feign psychiatric symptomatology to increase the likelihood they will receive a favorable legal outcome (Rogers, 2018). It was predicted that individuals with no external motivating factors, such as the inpatient sample, would be unlikely to overreport psychological impairment on the MMPI-3. The results support this hypothesis and suggest that evaluatees overreported a broad range of psychological, cognitive, and somatic symptoms not commonly seen in the community to a greater extent than inpatients. Interestingly, when examining the Fp scale, the results suggest that inpatients and evaluatees do not significantly differentially overreport psychiatric impairment or dysfunction. It is possible that, since the F scale captures a wider range of overreporting, the scale was more sensitive to modest overreporting than the Fp scale. Additionally, evaluatees may have been incentivized to overreport across multiple domains. Given the nature of a CST evaluation, it makes sense that evaluatees would be incentivized to overreport across different domains; CST requires cognitive skills (memory, reasoning) as well as psychiatric stability. In addition, given

the incentive to engage in response bias, it is conceivable that evaluatees attempted to overreport impairments broadly rather than regarding a specific domain of functioning.

As noted above, the Fp scale comprises infrequently endorsed symptoms by inpatients with genuine severe psychopathology, whereas the F scale comprises infrequently endorsed items by a community sample (Greene, 2011). Modest elevations on the F scale amongst inpatients with genuine severe psychiatric disorders are often indicative of genuine distress related to their mental illness. Importantly, clinically significant elevations were not found for the inpatient sample in the current study whereas clinically significant elevations were found for evaluatees. Overall, given the broader nature of the F scale, the results suggest that inpatients may have overreported psychological impairment/distress more broadly, indicating genuine impairments in psychological functioning.

Research has found that some items on the Fp scale are unlikely to be endorsed by forensic examinees under normal circumstances (Frederick, 1998). Further, Strong et al. (2006) found that the Fp scale successfully distinguished forensic evaluatees with high scores on the F scale who overreported severe psychopathology and those with genuine severe psychopathology. As such, evaluatees who display elevations on the F scale, but not the Fp scale may be displaying genuine impairment. This suggests that the forensic evaluatees in the current sample engaged in overreporting. As such, the results suggest that the evaluatee group engaged in some attempts to provide self-unfavorable reports, whereas this was not the case for inpatients. Ultimately, given that the majority of the evaluatee sample was ultimately found CST and/or criminally responsible, it is likely they were misrepresenting their abilities during their evaluations.

As described above, it is understandable that forensic evaluatees would have greater motivation to overreport psychological dysfunction and distress as well as experience greater

genuine distress than the inpatient sample. Whereas the outcome of the MMPI-3 had no bearing on clinical or legal outcomes for the inpatient samples, evaluatees were incentivized to portray psychological distress and impairment to potentially influence the outcomes of their overall forensic evaluations in a favorable way. After all, forensic evaluatees have been found to engage in various response biases to influence the outcomes of their evaluations (Melton et al., 2018).

In addition, the inflated scores (though not clinically significant) displayed by the inpatient sample may be due to genuine psychiatric impairment. Viljoen et al. (2002) suggests that inpatients are often treated for symptoms of severe mental illnesses, as well as other cognitive impairments. Although symptoms improve with treatment, some impairment often persists. As such, it is reasonable to predict that a group of inpatients with no external incentive would display greater genuine psychiatric impairment and distress than a community sample with no incentive. As such, the lack of differences observed between the groups on the Fp scale could be due to genuine psychological distress displayed by inpatients and attempts of evaluatees to engage in more broad and modest overreporting.

Hypothesis 4 examined whether evaluatees would experience differential motivation to present themselves as overly honest, moral, and conforming from inpatients, given evaluatees' vested interest in attaining a favorable evaluation result. Greene (2011) described that, in addition to denial of symptoms of psychopathology, the L scale also includes items assessing for the minimization of "minor, personal dishonesties and denial of aggression, bad thoughts, and weakness of character" (p. 79). It was hypothesized that evaluatees experienced unique pressures to present themselves not only as overly psychologically impaired and distressed, but also as more overly virtuous on the MMPI-3 with the idea that their presentation would contribute to a favorable evaluation outcome. Likewise, inpatients in the current study had no incentive to

portray themselves in an overly positive light or to otherwise manipulate their response styles on the MMPI-3. Available research suggests that forensic evaluatees would be unlikely to produce clinically significant elevations on the L scale, due to pressures to display genuine or exaggerated psychopathology. Indeed, Greene (2011) indicated that the L scale also assesses the minimization of psychological impairment and symptomatology in addition to uncommon virtues. However, it was anticipated that differential pressures experienced between forensic evaluatees and inpatients would contribute to statistically significant differences on the L scale due to differences in motivation to portray high virtuousness. Although there was limited research examining the L scale in criminal forensic evaluatees in comparison with forensic inpatients, the literature suggested that criminal defendants are capable of engaging in underreporting response bias when properly motivated (for example, see Bagbey et al., 1995). Therefore, it was anticipated that these differences would be evident through the L scale on the MMPI-3.

Despite the rationale for Hypothesis 4, the results suggest that evaluatees did not present themselves as differently honest, moral, and conforming when compared to inpatients. An examination of the mean values for the inpatient and evaluatee samples further suggests that both samples did not produce clinically significant elevations on the L scale, which was consistent with the hypothesis. As such, neither group presented themselves as overly virtuous. That suggests that evaluatees may not be as concerned with presenting themselves as highly virtuous as predicted. Consistent with prior research, the evaluatees instead engaged in overreporting, depicting significant distress and psychiatric impairment. It appears that individuals evaluated for CST and/or NGRI do not experience external pressure to present themselves in an overly positive light more so than non-incentivized inpatients. This finding suggests there is a stronger incentive for criminal forensic evaluatees to focus on displaying clinical impairment than

virtuousness. It is likely that since evaluatees in the current context often experience higher pressure to display psychiatric impairment and distress, an overreporting response style was the primary focus during the administration. However, it remains somewhat unclear whether evaluatees truly engaged in no attempt to manipulate their adherence to social norms and virtues. There are several alternative explanations for the current results.

The results could have been confounded by the incentive for evaluatees to overreport psychopathology and psychological distress, given that the L scale measures both uncommon virtues and denial of symptomatology (Greene, 2011). It is possible that a Type II error occurred; if the participants were administered a scale measuring only uncommon virtues and moral superiority, the groups may have displayed significant differences. In addition, Greene (2011) discussed that the L scale is also confounded by the face valid nature of some of the item content. Early studies on the MMPI found that the L scale was unsuccessful in detecting sophisticated respondents who were given instructions to underreport on the MMPI (see Vincent et al., 1966). As such, this sample of evaluatees may have had a level of sophistication, such that they recognized the nature of the items on the scale. Although Greene (2011) discussed that this level of sophistication is most frequently seen in groups with higher education and socioeconomic statuses, it is curious to consider factors that could impact sophistication amongst a sample of evaluatees. Indeed, most of the evaluatees in the sample were comprised of individuals who were ultimately opined competent to stand trial, which indicates that they had the ability to understand legal knowledge pertinent to their cases and also suggests they were capable of a higher degree of sophistication than a group of individuals opined incompetent. As such, it is possible that the sample also possessed a level of sophistication such as to alert them to the nature of some of the items on the L scale.

Despite the possible confounds, the results suggest that criminal forensic evaluatees are not likely to engage in a clinically significant effort to underreport psychiatric dysfunction or normal personal faults/weaknesses, nor are they likely to do so to a greater extent than individuals with no external incentives. Ultimately, the results of Hypothesis 4 indicate that problems of underreporting are unlikely in criminal forensic evaluations, particularly evaluations of competency to stand trial and criminal responsibility.

To further understand differences in response bias between inpatients and evaluatees, the remaining MMPI-3 validity scales were also examined. The results suggest that evaluatees are likely to overreport somatic (FBS) and cognitive (RBS) symptoms on the MMPI-3 to a greater extent than inpatients with no external incentive. An examination of the mean scores for evaluatees suggests that, as a group, they produced higher scores on FBS and RBS than inpatients. That suggests that motivating factors unique to the evaluation setting may incentivize evaluatees to overreport cognitive and somatic complaints. It is likely that evaluatees are motivated to present themselves as overly psychologically and/or physically impaired by a desire to positively influence the outcome of their evaluation and subsequent case, whereas inpatients in the current sample do not experience those same motivations. Further, it makes sense that inpatients displayed lower scores on the FBS and RBS validity scales, given they do not have external pressures to present themselves as overly psychologically or physically impaired.

The results also suggest that evaluatees are likely to present themselves as less well-adjusted and more open about their psychological problems than the inpatient sample as demonstrated by the difference observed between the two groups on the K scale. This suggests that evaluatees are likely incentivized to present their problems or even exaggerate their problems due to a desire to impact their evaluation and case in a desirable way. In addition, the inpatient

sample likely did not experience any motivation to approach the MMPI-3 defensively or overly openly. Given they were incentivized based upon their participation only rather than their performance, they had little reason to be overly forthcoming or defensive about their impairments and adjustment problems. Importantly, though significant differences were found, both samples provided average scores below the range of clinical significance. The mean scores suggest that neither inpatients ($M = 51.82$) nor evaluatees ($M = 45.24$) presented as overly defensive and well-adjusted. The interpretive manual for the MMPI-3 indicates that any score under T60 is indicative of no evidence of underreporting.

Conclusions

Overall, the results indicate that response bias on the MMPI-3 by individuals involved in the legal system can be influenced by specific forensic contexts. The current study examined differences and similarities exhibited by the different samples on consistency in responses, tendency to overreport symptoms, and tendency to underreport impairment.

Regarding response consistency, the expected different pressures and impairments between inpatients and evaluatees did not appear to differentially impact their response consistency on the MMPI-3. The results demonstrate that IST evaluatees and inpatients do not necessarily display higher rates of inconsistency or true response bias than NGRI evaluatees and inpatients. Further, the current study provided evidence that IST evaluatees and IST inpatients, as well as evaluatees and inpatients more broadly, do not necessarily differ in their tendency to respond consistently on the MMPI-3 despite differences in context. The current study also suggests that evaluatees are likely to overreport psychiatric impairment and distress to a greater extent than inpatients. Further, they are likely to produce clinically significant levels of overreporting in contrast to inpatients. Regarding underreporting symptoms and distress, the

current study suggests that though evaluatees are no more likely than inpatients to endorse uncommon virtues and psychological stability, they are likely to display greater openness regarding adjustment problems and a tendency to display themselves as less well-adjusted.

It is important to consider the broader implications of the current study. Scientists have long been fascinated with understanding external and underlying influences on behavior. Psychologists are tasked to not only listen to and interpret what individuals report, but also to understand the influences that might contribute to inaccurate reporting. Particularly in forensic contexts, it is important for psychologists to understand the different internal and external factors that can contribute to response bias. Fortunately, self-report psychological assessment instruments, such as the MMPI, have been useful tools in making important psycholegal decisions and opinions (Wygant & Granacher, 2015). Melton et al. (2018) highlighted that individuals within the legal process are likely to have differential motivating factors that could influence their reports during a psychological evaluation depending on their stage in the legal process as well as underlying psychological problems/deficits. For example, an individual who is at the start of the legal process may be motivated to overreport psychiatric or cognitive impairment such as to potentially avoid prosecution. Additionally, an individual evaluated for competency to stand trial may exhibit greater psychiatric impairment that influences their ability to take part in a psychological assessment than an inpatient who has received significant treatment. Understanding the nuances that influence response bias amongst different individuals within forensic contexts helps paint a clearer picture of what to expect in specific forensic contexts.

The results of the present study provide preliminary information that can be used to distinguish how response bias is similar and different between certain forensic groups. The most

significant difference was observed between evaluatees engaged in a legitimate forensic evaluation and inpatients who had no vested interest in the assessment. The results indicated that participation in a legitimate forensic evaluation was one of the most significant factors in influencing the tendency to engage in response bias, overreporting in particular. Understanding these distinctions is important for forensic healthcare providers. In one respect, the results can reduce stigma surrounding forensic evaluatees/patients. Specifically, the results indicate that forensic populations do not inherently engage in response bias simply by function of their involvement in the legal system. Rather, they are susceptible to influential factors such as mental health symptoms, cognitive impairment, and external motivating factors similarly to the rest of the population. Further, the current study suggests that commitment code, itself, may not be a defining characteristic that impacts response bias on the MMPI-3. It appears that despite apparent differences between IST inpatients and NGRI inpatients, both groups responded consistently on the MMPI-3 due to experiencing similar external pressures. Of course, further research is needed to determine whether these similarities exist between the same individuals earlier in the legal process.

This research serves as an important step in understanding the factors that influence response bias in forensic contexts. The distinctions and similarities between the groups examined provide clarity about what may contribute to certain response bias for specific patients or evaluatees. Ultimately, a more nuanced understanding of why certain individuals are susceptible to response bias will help psychologists produce more accurate and useful evaluations.

Limitations

Although several specific limitations of the current study were discussed in some detail in discussion of the results, it is important to highlight and consider the implications of such

limitations in further detail. There was considerable discrepancy in n between several of the sample groups. In particular, there were very few evaluatee participants who were opined IST or NGRI. As mentioned above, that difference could have impacted hypotheses one and two in particular. Both hypotheses originally aimed to identify differences between IST evaluatees and/or NGRI evaluatees and other groups. In addition, it was unclear whether the small samples of IST evaluatees and NGRI evaluatees were actually representative of the greater populations. As such, the results of related analyses may have limited internal and/or external validity.

Likewise, there was a relatively large discrepancy between the number inpatients and evaluatees utilized in the sample. Constraints on data collection led to a total of only 41 evaluatees, whereas the study included 114 inpatients. Given the limited sample of evaluatees, it remains unclear whether the sample is an accurate reflection of the evaluatee population, particularly given that the sample overwhelmingly consisted of evaluatees opined CST and/or criminally responsible.

Additionally, as noted above, the current sample of inpatients was fully aware that their participation was within the confines of a research study. They were also made aware that the results of their MMPI-3 would have no bearing on their legal or clinical status. As such, it is somewhat difficult to apply the results of their profiles to typical forensic assessment contexts. Although forensic inpatients are likely to be administered the MMPI-3 in a hospital setting, it is difficult to say how their results would compare to those of the current sample. The current inpatient sample had unique motivating factors that could have impacted their response styles, which are likely different from those experienced by other forensic inpatients prescribed the MMPI-3 as a component of a psychological assessment. The nature of the inpatient sample produces a threat to external validity.

This study employed a cut-off value of 70T on scales of non-content-based response bias when assembling the sample for analyses involving overreporting and underreporting. This is a more conservative criterion than is typically utilized in other studies regarding MMPI instruments. However, for purposes of this research, the authors sought to eliminate any impact of random responding on the results of hypotheses examining content-based validity scales. Since a *T* score of 70 is the minimum threshold indicating the possibility that non-content-based response bias impacted participant MMPI-3 scores, the authors were able to guarantee a sample free of non-content-based response bias. However, this strict criterion may have limited the sample size in the current study. Elevations (particularly minor elevations) in response bias do not necessarily invalidate a protocol, nor is there evidence to suggest minor elevations preclude the existence of simultaneous content-based distortions. As such, it might be worthwhile to implement a less conservative cut-off score in the future such as those outlined in the MMPI-3 technical manual.

Future Directions

The current study serves as an important step in better understanding how forensic populations approach the MMPI-3 differently when they experience different internal and external influences. The current study demonstrated that though different forensic evaluatees and inpatients display some similarities in their response biases, they also display important differences. Given the limitations outlined regarding the samples, it is recommended that future research examine response bias amongst a large sample of IST and NGRI evaluatees. As mentioned, there were fewer than 10 participants included in each evaluatee group for the current study. Additional research is needed to determine the extent to which IST and NGRI evaluatees display differences on MMPI-3 validity scales from other groups of evaluatees and inpatients. In

addition, given current trends in examining research reliability, it will be important to replicate the findings of the current study before drawing definitive conclusions from the results. Such additional and validating research could provide important information about the factors that impact profile validity for these populations.

In addition, future research should also work to distinguish the role of motivational factors and psychological differences in the response styles observed. For nearly each finding, possible explanations for certain response styles included both external motivating factors as well as psychological differences. For example, it was hypothesized that differences in overreporting between evaluatees and inpatients could have resulted due to differences in situational pressures as well as differences in psychiatric functioning. It remains to be researched the extent of the impact of each on response bias. In an effort to provide some direction for future research, it is recommended that future studies examine participants with higher scores on clinical scales and/or those who display greater psychopathology. An examination of individuals who display elevations on certain clinical scales could give information about how certain symptoms of mental illness or psychological impairment uniquely influence response bias. Further, it remains unclear whether IST and NGRI inpatients provided similar scores of inconsistency due to psychological or motivational similarities. It would be useful for future research to parse apart the differential impacts of external motivating factors and psychological differences on response bias in forensic populations.

The current study provides initial data regarding important differences between forensic subpopulations that impact response bias. These differences can inform practitioners and researchers about psychological and motivational differences that impact how people are likely to approach a psychological assessment. Further exploration of this topic will better inform

researchers and practitioners about how to best minimize the impacts of different response biases in high-stakes forensic assessments and evaluations.

References

- Arbisi, P. A., & Ben-Porath, Y. S. (1998). The ability of Minnesota Multiphasic Personality Inventory-2 validity scales to detect fake-bad responses in psychiatric inpatients. *Psychological Assessment, 10*(3), 221–228.
- Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment, 14*, 16–26.
- Bagby, R. M., Rogers, R., Nicholson, R. A., Buis, T., Seeman, M. V., & Rector, N. A. (1997). Effectiveness of the MMPI-2 validity indicators in the detection of defensive responding in clinical and nonclinical samples. *Psychological Assessment, 9*(4), 406–413.
- Ben-Porath, Y. S. (2012a). Addressing challenges to MMPI-2-RF-based testimony: Questions and answers. *Archives of Clinical Neuropsychology, 27*, 691–705.
- Ben-Porath, Y. S. (2012b). *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. University of Minnesota Press.
- Berman, L. M., & Osborne, Y. H. (1987). Attorneys' referrals for competency to stand trial evaluations: Comparisons of referred and nonreferred clients. *Behavioral Sciences & the Law, 5*(3), 373–380.
- Bernreuter, R. G. (1935). *Manual for the Personality Inventory*. Stanford University Press.
- Boccaccini, M. T., Murrie, D. C., & Duncan, S. A. (2006). Screening for malingering in a criminal-forensic sample with the Personality Assessment Inventory. *Psychological Assessment, 18*, 415–423.

- Bow, J. N., Flens, J. R., & Gould, J. W. (2010). MMPI-2 and MCMI-III in forensic evaluations: A survey of psychologists. *Journal of Forensic Psychology Practice, 10*, 37–52.
- Butcher, J. N. (2000). Revising psychological tests: Lessons learned from the revision of the MMPI. *Psychological Assessment, 12*(3), 263-271.
- Butcher, J. N., Arbisi, P. A., Atlis, M. M., & McNulty, J. L. (2003). The construct validity of the Lees-Haley Fake Bad Scale: Does this scale 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 restandardized Minnesota Multiphasic Personality Inventory: MMPI-2. An administrative and interpretive guide*. University of Minnesota Press.
- Butcher, J. N., Graham, J. R., Williams, C. L., & Ben-Porath, Y. (1990). *Development and use of the MMPI-2 content scales*. University of Minnesota Press.
- Butcher, J. N. & Hostetler, K. (1990). Abbreviating MMPI item administration: What can be learned from the MMPI for the MMPI-2? *Psychological Assessment, 2*, 12-21.
- Carpenter, W. T., Jr., Gold, J. M., Lahti, A. C., Queern, C. A., Conley, R. R., Bartko, J. J., Kovnick, J., & Appelbaum, P. S. (2000). Decisional capacity for informed consent in schizophrenia research. *Archives of General Psychiatry, 57*(6), 533-538.
- Clopton, J. R., & Neuringer, C. (1977). MMPI cannot say scores: Normative data and degree of profile distortion. *Journal of Personality Assessment, 41*(5), 511-513.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1975). *An MMPI handbook, Vol. 2: Research applications*. University of Minnesota Press.
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- Durham v. United States, 214 F.2d 862 (1954).

- Fjordbak, T. (1985). Clinical correlates of high lie scale elevations among forensic patients. *Journal of Personality Assessment, 49*(3), 252-255.
- Frederick, R. I. (1998). *Characteristics of the Infrequency-Psychopathology scale, Fp, among a sample of criminal forensic examinees*. [Conference presentation]. APA 1998 Convention, San Francisco, CA, United States.
- Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior, 24*, 59–82.
- Friedman, A. F., Bolinsky, P. K., Levak, R. W., & Nichols, D. S. (2015). *Psychological assessment with the MMPI-2/MMPI-2-RF* (3rd ed.). Routledge.
- 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.
- Gibby, R. E., & Zickar, M. J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology, 11*(3), 164-184.
- Glassmire, D. M., Jhavar, A., Burchett, D., & Tarescavage, A. M. (2017). Evaluating item endorsement rates for the MMPI-2-RF F-r and fp-r scales across ethnic, gender, and diagnostic groups with a forensic inpatient sample. *Psychological Assessment, 29*(5), 500-508.
- Graham, J. R. (2012). *MMPI-2: Assessing personality and psychopathology* (5th ed.). Oxford University Press.
- Graham, J. R., Watts, D., & Timbrook, R. E. (1991). Detecting fake good and fake bad MMPI-2 profiles. *Journal of Personality Assessment, 57*, 264-277

- Greene, R. L. (2007). Forensic applications of the Minnesota Multiphasic Personality Inventory-2. In A. M. Goldstein (Ed.), *Forensic psychology: Emerging topics and expanding roles* (pp. 73-96). John Wiley & Sons Inc.
- Greene, R. L. (2011). *The MMPI-2/MMPI-2-RF: An interpretive manual* (3rd ed.). Allyn & Bacon.
- Grossi, L. M., Green, D., Einzig, S., & Belfi, B. (2017). Evaluation of the response bias scale and improbable failure scale in assessing feigned cognitive impairment. *Psychological Assessment, 29*(5), 531-541.
- Groth-Marnat, G. (2009). *Handbook of psychological assessment*. John Wiley & Sons.
- Gu, W., Reddy, H. B., Green, D., Belfi, B., & Einzig, S. (2017). Inconsistent responding in a criminal forensic setting: An evaluation of the VRIN-r and TRIN-r scales of the MMPI-2-RF. *Journal of Personality Assessment, 99*(3), 286-296.
- Hale, N. G., Jr. (1995). *The rise and crisis of psychoanalysis in the United States*. Oxford University Press.
- Hall, J., Menton, W., & Ben-Porath, Y. S. (2020, June). *Examining the psychometric equivalency of MMPI-3 scale scores derived from the MMPI-3 and the MMPI-2-RF-EX*. [Conference presentation]. 55th Annual Symposium on Recent MMPI Research, Minneapolis, MN, United States.
- Hall, J. T., Menton, W. H., & Ben-Porath, Y. S. (2021). Examining the psychometric equivalency of MMPI-3 scale scores derived from the MMPI-3 and the MMPI-2-RF-EX. *Assessment*. Advance online publication.

- Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *The Journal of Psychology: Interdisciplinary and Applied*, *10*, 249–254.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Schedule*. University of Minnesota Press.
- Herbel, B. L., & Stelmach, H. (2007). Involuntary medication treatment for competency restoration of 22 defendants with delusional disorder. *Journal of the American Academy of Psychiatry and the Law*, *35*(1), 47-59.
- Himsl, K., Burchett, D., Tarescavage, A. M., & Glassmire, D. M. (2017). Assessing reading ability for psychological testing in forensic assessments: An investigation with the WRAT-4 and MMPI-2-RF. *The International Journal of Forensic Mental Health*, *16*(3), 239-248.
- Hoge, S. K., Bonnie, R. J., Poythress, N., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1997). The MacArthur adjudicative competence study: Development and validation of a research instrument. *Law and Human Behavior*, *21*(2), 141-179.
- Humm, D. G., & Wadsworth, G. W., Jr. (1934). The Humm-Wadsworth Temperament Scale: Preliminary report. *Personnel Journal*, *12*, 314.
- Huprich, S. K., & Ganellen, R. J. (2006). The advantages of assessing personality disorders with the Rorschach. In S. K. Huprich (Ed.), *Rorschach assessment to the personality disorders* (pp. 27-53). Lawrence Erlbaum Associates Publishers.
- 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*, *30*(4), 473-496.

- James, D. V., Duffield, G., Blizard, R., & Hamilton, L. W. (2001). Fitness to plead: A prospective study of the inter-relationships between expert opinion, legal criteria and specific symptomatology. *Psychological Medicine, 31*, 139–150.
- Kucharski, L. T., Johnsen, D., & Procell, S. (2004). The utility of the MMPI-2 infrequency psychopathology F(p) and the revised infrequency psychopathology scales in the detection of malingering. *American Journal of Forensic Psychology, 22*(1), 33-40.
- 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*(1), 203–210.
- LePage, J. P., & Mogge, N. L. (2001). Validity rates of the MMPI-2 and PAI in a rural inpatient psychiatric facility. *Assessment, 8*(1), 67-74.
- Locke, D. E. C., Kirilin, K. A., Thomas, M. L., Osborne, D., Hurst, D. F., Draskowski, J. F., Servin, J. I., & Noe, K. H. (2010). The Minnesota Multiphasic Personality Inventory-2-restructured form in the epilepsy monitoring unit. *Epilepsy & Behavior, 17*(2), 252-258.
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist, 29*, 741–776.
- Matlasz, T. M., Brylski, J. L., Leidenfrost, C. M., Scalco, M., Sinclair, S. J., Schoelerman, R. M., Tsang, V., & Antonius, D. (2017). Cognitive status and profile validity on the personality assessment inventory (PAI) in offenders with serious mental illness. *International Journal of Law and Psychiatry, 50*, 38-44.
- McCusker, P. J., Moran, M. J., Serfass, L., & Peterson, K. H. (2003). Comparability of the MMPI-2 F(p) and F scales and the SIRS in clinical use with suspected

- malingers. *International Journal of Offender Therapy and Comparative Criminology*, 47(5), 585-596.
- Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, 30(5), 525-564.
- Melton, G. B., Petrila, J., Poythress, N. G., Slobogin, C., Otto, R. K., Mossman, D., Condie, L. O. (2018). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (4th ed.) Guilford Press.
- Mendoza, R. J. (2005). The need for involuntary medication in competency to stand trial proceedings: A necessary work in progress. *Journal of Forensic Psychology Practice*, 5(3), 69-78.
- Morse, S. J., & Morse, S. J. (1980). Beating a rap? Defendants found incompetent to stand trial. *Sociology and Social Research*, 64(4), 574-576.
- Moskowitz, J. L., Lewis, R. J., Ito, M. S., & Ehrmentraut, J. (1999). MMPI-2 profiles of NGRI and civil patients. *Journal of Clinical Psychology*, 55(5), 659-668.
- Neal, T. M., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior*, 41(12), 1406–1421.
- Nestor, P. G., Daggett, D., Haycock, J., & Price, M. (1999). Competence to stand trial: A neuropsychological inquiry. *Law and Human Behavior*, 23(4), 397-412.
- Nestor, P. G., & Haycock, J. (1997). Not guilty by reason of insanity of murder: Clinical and neuropsychological characteristics. *Journal of the American Academy of Psychiatry and the Law*, 25(2), 161-171.

- Nicholson, R. A., Briggs, S. R., & Robertson, H. C. (1988). Instruments for assessing competency to stand trial: How do they work? *Professional Psychology: Research and Practice*, *19*(4), 383-394.
- Pancoast, D. L., & Archer, R. P. (1989). Original adult MMPI norms in normal samples: A review with implications for future developments. *Journal of Personality Assessment*, *53*(2), 376-395.
- Pirelli, G., Gottdiener, W. H., & Zapf, P. A. (2011). A meta-analytic review of competency to stand trial research. *Psychology, Public Policy, and Law*, *17*(1), 1-53.
- Queen v. M'Naghten, 8 Eng. Rep. 718 (1843).
- Ray, C. L. (2017). Practical use of MMPI-2-RF validity indicators in VA compensation and pension examinations. *Psychological Injury and Law*, *10*(3), 223-233.
- Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, *10*, 10-20.
- Rey, A. (1964). *L'examen clinique en psychologie [Clinical examination in psychology]*. Presses Universitaires de France.
- Richey, L. N., & Doninger, N. A. (2020). A comparison of performance validity measures in predicting MMPI-2 lie scale results. *The Clinical Neuropsychologist*, *34*(2), 353-367.
- Rogers, R. (2008). *Clinical assessment of malingering and deception* (3rd ed.). Guilford Press.
- Rogers, R. (2018). An introduction to response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (4th ed., pp. 3-18) Guilford Press.
- Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured interview of 30 reported symptoms: Professional manual*. Psychological Assessment Resources.

- 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.
- Rosanoff, A. J. (1920). A theory of personality based mainly on psychiatric experience. *Psychological Bulletin, 17*, 281–299.
- Samuel, S. E., DeGirolamo, J., Michals, T. J., & O'Brien, J. (1994). Preliminary findings on MMPI "cannot say" responses with personal injury litigants. *American Journal of Forensic Psychology, 12*(4), 5-18.
- Sánchez, G., Ampudia, A., Jiménez, F., & Amado, B. (2017). Contrasting the efficacy of the MMPI-2-RF overreporting scales in the detection of malingering. *The European Journal of Psychology Applied to Legal Context, 9*(2), 51-56.
- Schroeder, R. W., Baade, L. E., Peck, C. P., VonDran, E. J., Brockman, C. J., Webster, B. K., & Heinrichs, R. J. (2012). Validation of MMPI-2-RF validity scales in criterion group neuropsychological samples. *The Clinical Neuropsychologist, 26*(1), 129-146.
- Sellbom, M. (2017). Using the MMPI-2-RF to characterize defendants evaluated for competency to stand trial and criminal responsibility. *The International Journal of Forensic Mental Health, 16*(4), 304-312.
- Sellbom, M., & Bagby, R. M. (2008). Validity of the MMPI-2-RF (restructured form) L-r and K-r scales in detecting underreporting in clinical and nonclinical samples. *Psychological Assessment, 20*(4), 370-376.
- Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T., & Duncan, S. (2010). Utility of the MMPI-2-RF (restructured form) validity scales in detecting malingering in a criminal forensic setting: A known-groups design. *Psychological Assessment, 22*(1), 22-31.

- Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, *15*, 201–293.
- Strong, D. R., Glassmire, D. M., Frederick, R. I., & Greene, R. L. (2006). Evaluating the latent structure of the MMPI-2 Fp scale in a forensic sample: A taxometric analysis. *Psychological Assessment*, *18*, 250-261.
- Super, D. E. (1942). The Bernreuter Personality Inventory: A review of research. *Psychological Bulletin*, *39*, 94–125.
- Super, D. E. (1949). Review of assessment of men: Selection of personnel for the office of strategic services. *Journal of Applied Psychology*, *33*(5), 511-515.
- Toomey J. A., Kucharski, L. T., & Duncan, S. (2009). The utility of the MMPI-2 malingering discriminant function index in the detection of malingering: A study of criminal defendants. *Assessment*, *16*(1), 115-121.
- Ustad, K. L., Rogers, R., Sewell, K. W., & Guarnaccia, C. A. (1996). Restoration of competency to stand trial: Assessment with the georgia court competency test and the competency screening test. *Law and Human Behavior*, *20*(2), 131-146.
- Vallabhajosula, B., & van Gorp, W. G. (2001). Post-Daubert admissibility of scientific evidence on malingering of cognitive deficits. *Journal of the American Academy of Psychiatry and the Law*, *29*, 207–215.
- Viljoen, J. L., Roesch, R., & Zapf, P. A. (2002). An examination of the relationship between competency to stand trial, competency to waive interrogation rights, and psychopathology. *Law and Human Behavior*, *26*(5), 481-506.
- Vincent, N. M., Linsz, N. L., & Greene, M. I. (1966). The L scale of the MMPI as an index of falsification. *Journal of Clinical Psychology*, *22*(2), 214-215.

- Warren, J. I., Fitch, W. L., Dietz, P. E., & Rosenfeld, B. D. (1991). Criminal offense, psychiatric diagnosis, and psycholegal opinion: An analysis of 894 pretrial referrals. *Bulletin of the American Academy of Psychiatry & the Law*, *19*(1), 63-69.
- Wasyliv, O. E., Grossman, L. S., Haywood, T. W., & Cavanaugh, J. L. (1988). The detection of malingering in criminal forensic groups: MMPI validity scales. *Journal of Personality Assessment*, *52*(2), 321-333.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Addison-Wesley Publishing Co.
- Woodworth, R. S. (1917). *Personal data sheet*. C. H. Stoelting Company.
- Woodworth, R. S. (1919). Examination of emotional fitness for warfare. *Psychological Bulletin*, *16*, 59–60.
- Wygant, D. B., Ben-Porath, Y. S., & Arbisi, P. A. (2004, May). *Development and initial validation of a scale to detect infrequent somatic complaints*. [Conference presentation]. 39th Annual Symposium on Recent Developments of the MMPI-2/MMPI-2-A, Minneapolis, MN, United States.
- 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*(7), 671-680.
- Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology*, *22*, 489 – 499.

- 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*, 745–756.
- Wygant, D. B., Walls, B. D., Brothers, S. L., & Berry, D. T. R. (2018). Assessment of malingering and defensiveness on the MMPI-2 257 and MMPI-2-RF. In R. Rogers, & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 257-279). The Guilford Press.