Development of a Correction for Defensiveness in the Minnesota Multiphasic Personality Inventory-Second Edition-Restructured Form (MMPI-2-RF) Profiles of Adult Male Sex Offenders

by

Katie Marie Glauner, M.A., M.S.

Master of Science
in Clinical Psychology
Florida Institute of Technology
2020

Master of Arts
in Forensic Psychology
University of Denver
2018

Bachelor of Science
in Psychology and Criminal Justice & Criminology
Arizona State University
2015

A doctoral research project submitted to
The College of Psychology and Liberal Arts at
Florida Institute of Technology
in partial fulfilment of the requirement
for the degree of

Doctor of Psychology
in
Clinical Psychology

Melbourne, Florida
October 2021

The undersigned committee hereby recommends that the attached document be accepted as fulfilling in part the requirements for the degree of Doctor of Psychology in Clinical Psychology

Development of a Correction for Defensiveness in the Minnesota Multiphasic Personality Inventory-Second Edition-Restructured Form (MMPI-2-RF) Profiles of Adult Male Sex Offenders

by

Katie Marie Glauner, M.A., M.S.

_____

Radhika Krishnamurthy, Psy.D.
Professor
School of Psychology
Major Advisor

_____

Julie Costopoulos, Ph.D.
Associate Professor
School of Psychology

_____

Ashok Pandit, Ph.D.
Professor
Mechanical and Civil Engineering

_____

Robert A. Taylor, Ph.D.
Professor and Interim Dean
College of Psychology and Liberal Arts

Abstract

Development of a Correction for Defensiveness in the Minnesota Multiphasic Personality Inventory-Second Edition-Restructured Form (MMPI-2-RF) Profiles of Adult Male Sex Offenders

by

Katie Marie Glauner, M.A., M.S.

Committee Chair: Radhika Krishnamurthy, Psy.D., ABAP

Clinical and forensic psychologists often evaluate sex offenders to determine the level of risk they pose to the community and identify their treatment needs to reduce the risk of future recidivism. Personality assessments are administered in sex offender evaluations to aid in answering these referral questions, while also providing information regarding the evaluees' response style to the test and descriptions of their personality functioning. The Minnesota Multiphasic Personality Inventory (MMPI) and its revisions have been the most widely used personality assessments in sex offender evaluations, with previous research demonstrating that sex offenders often respond defensively to the test by minimizing or denying their psychological problems, thus limiting the interpretability of the test results. The current study aimed to develop empirically-derived optimal cutting scores for various Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) substantive scales (i.e., Higher-Order (H-O), Restructured Clinical (RC), Specific Problems (SP), and Personality Psychopathology Five (PSY-5) scales) for sex offenders to adjust for defensiveness and denial of psychopathology. Archival MMPI-2-RF data from a sample of $N = 142$ adult male sex offenders, previously deemed a defensive subgroup through cluster analysis, was compared to MMPI-2-RF data collected from a community sample of $N = 135$ adult men to compare means and standard

deviations on the substantive scales and derive optimal cutting scores for sex offenders using receiver operating characteristic (ROC) analyses. Multivariate analysis of variance (MANOVA) results followed by a series of univariate analyses of variance (ANOVAs) demonstrated statistically significant differences in scores between the sex offender sample and the community comparison sample on 29 of the 40 substantive scales. ROC analyses produced area under the curve (AUC) values of greater than .70 for three of the substantive scales (i.e., RC7, RC9, COG). With the exception of the RC6 and JCP scales that produced very low AUC values and exceptionally high optimal cutting scores, the optimal cutting scores for all other substantive scales fell between 40.5 (SHY) and 62.5 (AGGR-r). Specifically, three scales fell between optimal cutting scores of 40-44, 14 scales between 44-49, 13 scales between 50-54, seven scales between 55-59, and one scale between 60-64. Alternative cutting scores at equal intervals of 40, 45, 50, and 55 showed that a T score of 45 was optimal for 10 scales, T 50 for 25 scales, and T 55 for five scales. These empirically-derived optimal cutting scores and alternative cutting scores developed to enhance practical applications can potentially be used in sex offender evaluations to adjust for defensive responding, providing a more accurate interpretation of sex offenders' personality characteristics and psychopathology. Implications of these findings were discussed.

Table of Contents

List of Tables

Chapter 1: Introduction

Sex offenses present a significant concern to society. According to the National Center for Missing and Exploited Children (2016), there are more than 850,000 registered sex offenders in the United States. However, this number is not an accurate representation of the total number of sex offenders as it is limited only to those who have been convicted of a sex offense and required to register as a sex offender. An accurate estimate of the number of sex offenders in the United States is difficult to obtain as sex offenses are one of the most under-reported crimes. Thus, many offenders will never be detected, arrested, convicted, or registered for having committed a sex offense (Cubellis et al., 2019). What is known is that sexual victimization is far more common than what current sources report (Wiseman, 2015).

Individuals across all ages, races, ethnicities, and sexual orientations have been victims of sexual abuse. Sexual abuse as defined by the United States Department of Justice (n.d.) includes "any sexual act committed against someone without that person's freely given consent" that includes both touching offenses (e.g., forced or attempted sexual intercourse, child molestation, groping) and nontouching offenses (e.g., downloading or distributing child pornography, exhibitionism). It is important to note that the legal definitions of sex offenses differ between states and at the federal level. In the state of Florida, sex offenses are classified into three primary categories including Lewd or Lascivious Acts, Sexual Battery and Rape, and Unlawful Sex with Minors. Penalties for engaging in sex offenses in Florida can range from probation and incarceration to civil commitment.

Laws that have been enacted as punishment for committing sex offenses have become more extensive and restrictive over the last few decades, likely due to increased media reports on sex offenders and the sensational headlines of sex crimes that have come forth (Cubellis et al., 2015; Cucolo & Perlin, 2013). With advances in technology and mass media, there is a constant inundation of information on recently committed sex offenses that make it seem as if sex offenses are increasing exponentially in frequency. This influx of media attention, especially when highlighting offenses against children, provoke fear and panic within communities (Cucolo & Perlin, 2013). As such, conspiracy theories and conspiracy groups such as QAnon have emerged in the United States. QAnon, established in 2017, purports that America is being run by a cabal of pedophiles who run a global child-sex trafficking operation. This group has since emerged into mainstream media in 2020, perpetuating misinformation on pedophilia and leading to public uproar while inciting violence against suspected pedophiles (Wendling, 2020).

Sex offenders are one of the most highly stigmatized groups in society and can be identified as such by the punishments they face once convicted of a sex offense. The most common policies that have been enacted in recent years include sex offender registration, community notification, residency restrictions, and civil commitment (Levenson & D' Amora, 2007). In 1994, U.S. Congress passed the Jacob Wetterling Crimes Against Children and Sexually Violent Offender Registration Act that required all states to create laws that mandate sex offenders to register their personal information with their local law enforcement agency. These registries typically include the offender's home address, physical characteristics, crime offense, and sometimes vehicle information, making these offenders easily identifiable to others in the community. Shortly after, Megan's Law was

enacted in 1996 that required all states to develop a community notification system that distributes information to the public about registered sex offenders who live in close proximity. In addition to these federal laws, many states have developed additional policies such as residency restrictions in which registered sex offenders are prohibited from living within close proximity of schools, parks, day care centers, playgrounds, and other places that children frequent. Specifically, in Florida sex offenders are banned from occupying hurricane shelters and homeless shelters. In some ordinances, landlords can face criminal penalties if they knowingly rent their property to a registered sex offender (Levenson & D'Amora, 2007). These residency restrictions often make it difficult for sex offenders to find adequate housing and employment. Currently, 20 states have enacted laws that sanction civil commitment of sex offenders if they qualify as sexually dangerous post incarceration or conviction (Cucolo & Perlin, 2013). This involuntary commitment removes sexually dangerous offenders from society for an extended, sometimes indefinite, period of time. These federal and state level policies allot sex offenders with very little privacy and subjects them to a high level of scrutiny from community members.

Sex offenders experience stigmatization in a multitude of ways, ranging from discrimination and ostracism from social participation to physical assault and murder. Registered sex offenders are at an increased risk of experiencing harassment and victimization due to their placement on registries, with one study finding that between 5 and 16% of sex offenders have experienced physical assault in their lifetime (Levenson & Cotter, 2005). Research has found that civilians tend to support vigilantism against sex

offenders due to the perceived extreme nature of the crimes that sex offenders commit (Cubellis et al., 2019).

Given the heightened media attention, stigmatization, severe restrictions, and punishments ranging from probation and incarceration to civil commitment, it comes as no surprise that sex offenders are often defensive and engage in denial (Haywood et al., 1993). These characteristics are also evident in how sex offenders typically present themselves while undergoing sex offender evaluations. Sex offenders may undergo an evaluation for a variety of reasons, including to assess their level of dangerousness, determine their risk of recidivism, and to identify their treatment needs. These evaluations typically include a personality assessment, such as the Minnesota Multiphasic Personality Inventory-Second Edition-Restructured Form (MMPI-2-RF), that can assess an evaluee's response style and personality functioning. Previous research has found that sex offenders have a tendency to respond defensively to the MMPI and its subsequent revisions, meaning they minimize or deny their psychological symptoms in order to appear more psychologically well-adjusted (Haywood et al., 1994; Tarascavage et al., 2018). However, this defensive response style interferes with detecting an evaluee's true maladjustment, limiting the interpretability or even invalidating the test results altogether. The current study aimed to develop empirically-derived optimal cutting scores for MMPI-2-RF profiles of sex offenders to adjust for defensive responding. These optimal cutting scores will assist evaluators in providing a more accurate interpretation of sex offenders' psychological adjustment. This is vital considering the high-stakes decisions that are made using information gathered from sex offender evaluations.

Chapter 2: Review of the Literature

Sex Offender Characteristics

Researchers have attempted to understand the etiology of sex offending through the exploration of various psychological theories and the study of shared characteristics among sex offender populations. Numerous theories have been developed describing various biological, psychological, and social factors that may predispose someone to sex offending. Early psychoanalytic theories speculated that sexual deviancy arises from unresolved conflicts experienced during an individual's early stages of development (Wood et al., 2000). However, these theories have largely been replaced with more recent developmental theories such as attachment theory. Attachment theory speculates that sex offenders had poor quality attachments with their primary caretakers during their early developmental years and this led to interpersonal deficits in adulthood. Additionally, other cognitive, behavioral, and social learning theories have emerged to explain the etiology of sex offending as this issue is too complex and multifaceted to explain solely from a single theory (Faupel & Przybylski, 2015).

Research has shown that sex offenders are a heterogeneous group. The sex offender population varies in age, background, personality features, psychiatric diagnosis, life style, and types of victims they perpetrate against (Bard et al., 1987). In an effort to specify more homogenous subgroups of offenders, researchers have identified typologies of sex offenders.

Sex offender typologies are predominantly based on the age of the victim and the type of sexual offense committed (Simons, 2015). Therefore, many research studies evaluate the characteristics between contact versus non-contact sex offenses and

characteristics of child sexual abusers versus adult sexual abusers. Samples of sex offenders are frequently characterized by demographic information, by victim characteristics, by psychiatric diagnosis, and by personality features (Wood et al., 2000). These typologies were developed in an attempt to identify individuals who are at risk of committing sexual offenses and to identify treatment and supervision needs of these offenders (Polaschek et al., 1997).

Some of the most well-known and researched types of sex offenders include child sexual abusers, rapists, and internet sexual offenders (Beech et al., 2008). Empirical studies have been conducted to assess for the similarities and differences in characteristics between these sex offender typologies. Regarding demographic characteristics, child sexual abusers have been one of the most difficult groups to classify as they vary widely in socioeconomic status, marital status, ethnicity, and sexual orientation (Simons, 2015). However, the majority of all sex offenses are committed by adult men (Beech et al., 2008). Regarding educational attainment, child sexual abusers are typically less educated than internet sexual offenders and rapists (Bard et al., 1987; Laws & O'Donohue, 2008). Compared to rapists, child sexual abusers are less likely to have engaged in intimate relationships, to be married, or to have cohabitated with a partner for at least one year (Simons, 2015; Bard et al., 1987). The developmental histories of child sexual abusers have also been examined. Research suggests that child sexual abusers, specifically pedophiles, report experiencing higher rates of sexual abuse or trauma during their childhood compared to sex offenders who perpetrate against adults (Phenix & Hoberman, 2016). Regarding comorbidity with other psychiatric diagnoses, sex offenders who have been diagnosed with pedophilia are often also diagnosed with

other Diagnostic and Statistical Manual of Mental Disorders (DSM) Axis I (clinical) and Axis II (personality) disorders (Phenix & Hoberman, 2016). Other comorbidity studies have shown high levels of psychopathology among pedophiles, such as depression and anxiety disorders and personality problems (Laws & O'Donohue, 2008).

Empirical studies conducted on child sexual abusers have found that this population often demonstrates poor social skills, low self-esteem, and feelings of loneliness, and are also nonassertive in relationships (Simons, 2015; Laulik et al., 2007). Some researchers have speculated that child sexual offenders perpetrate in order to reduce their loneliness, depression, and anxiety symptoms (Simons, 2015). Compared to other sex offender typologies, child sexual abusers differ in their thought processes and utilization of cognitive distortions. Child sexual offenders often use cognitive distortions to minimize, rationalize, or justify their criminal behavior. Compared to other sex offender groups, these offenders were more likely to view child-adult sexual activity as socially acceptable, to believe children desired to have sex with adults, and that children are unharmed from sexual engagement with adults (Hayashino et al., 1995). Additionally, when compared to rapists, child sexual offenders demonstrated a greater fear of negative evaluation from others, suggesting this population is more sensitive to rejection and criticism from adults (Hayashino et al., 1995).

In comparison to child sexual abusers' sociodemographic characteristics, rapists are generally younger, tend to have lower socioeconomic status, are more likely to have had an intimate relationship, are less passive in relationships, and are more socially competent (Laws & O'Donohue, 2008). In evaluating the developmental histories of

rapists, this group reported more frequent physical abuse, emotional abuse, and parental violence during childhood in comparison to child sexual abusers (Simons, 2015).

Various studies have been conducted evaluating the psychiatric diagnoses among rapists. Studies have found that rapists are more likely to be diagnosed with a personality disorder, a substance use disorder, or psychosis than child sexual abusers (Langstrom et al., 2004; Craissati, 2005). Additionally, rapists are more likely to be diagnosed with psychopathy, or to exhibit psychopathic traits, than any other type of sex offender (Phenix & Hoberman, 2016). Rapists have been characterized as having intimacy deficits, sexual and general self-regulation deficits, negative peer influences, and adversarial attitudes or beliefs in support of their violent behaviors (Craissati, 2005). Additionally, research has shown a consistent positive relationship between sexually aggressive behaviors and the acceptance of rape myths, traditional sex roles, and negative attitudes about women (Polaschek et al., 1997).

A meta-analysis conducted by Whitaker et al. (2008) found that sex offenders who perpetrated against adults showed significantly higher levels of externalizing behaviors (e.g., violence, delinquency) compared to sex offenders who perpetrated against children. These offenders have also been found to have more previous convictions for violent crimes and are more likely to demonstrate greater force and aggression while committing their sexual offenses (Bard et al., 1987). Rapists are more likely to reoffend by committing non-sexual violent crimes than to commit additional sex crimes and closely resemble violent offenders in terms of personality characteristics. In examining why individuals sexually offend against adults, the most commonly reported motives for rape reported by offenders are those of anger, power, and sex (McCabe &

Wauchope, 2005). Studies conducted on sex offender samples have found that rapists frequently report experiencing anger prior to committing their sexual offense (Ramirez et al., 2015).

Numerous studies have examined sex offender characteristics utilizing personality measures in an attempt to find a common personality profile among various sex offender typologies. Armentrout and Hauer (1978) collected and analyzed Minnesota Multiphasic Personality Inventory (MMPI) results from a sample of 13 rapists, 21 child sexual abusers, and 17 individuals with non-rape offenses. The results of the study found that the rapist sample produced an elevated 8-4 profile (i.e., elevated scores on Schizophrenia and Psychopathic Deviate scales) while the child sexual abuser sample produced a less elevated but prominent 4-8 profile, suggesting anger, distrust, alienation, and impulsivity characteristics in both groups. However, rapists presented as more angry, alienated, and resentful than the child sexual abusers. Similar results were shown in a study conducted by Hall et al. (1986), finding that clinical scales 4 and 8 were significantly elevated in the mean MMPI profile of a child sexual abuser sample (n = 406). A 4-8 profile was found to be the overall mean code type; however, this elevation was found in less than 10% of the offender profiles and there was no single two-point code type that was prominent among the sample. Erickson et al. (1987) found that in administering the MMPI to a large sample (N = 403) of rapists and child sexual abusers, scale 4 was the only high scale that was prevalent in the majority of sex offender profiles. The 4-8/8-4 profiles were found to be the most common code types among the entire sex offender sample. Using an alternative personality measure, Chantry and Craig (1994) administered the Millon Clinical Multiaxial Personality Assessment (MCMI) to a group of violent offenders that

included rapists (*n* = 195), child sexual abusers (*n* = 201), and non-sexually aggressive felons (*n* = 205). Compared to the non-sexually aggressive felons, the two sex offender groups demonstrated more passive-aggressive features; however, the child sexual abusers appeared more passive-aggressive, dependent, depressed, and anxious and less self-focused, conforming, and suspicious than the rapists and non-sexually aggressive felon samples. Additionally, the child sexual abusers presented as more passive, insecure, and lacking in initiative while rapists and non-sexually aggressive felons presented as more narcissistic, independent, entitled, and less psychiatrically distressed. Similarly, Ahlmeyer and colleagues (2003) analyzed MCMI-III results of 223 rapists, 472 child sexual abusers, and 7,226 non-sexual offenders. Results found that the non-sexual offenders presented with more antisocial, narcissistic, and sadistic patterns, consistent with "classic" criminal personality styles, while the sex offender samples demonstrated broader and more severe psychopathology. Compared to the rapist sample, the child sexual abuser sample appeared more neurotic, affective, and socially impaired. These results are largely comparable to the results found in Chantry and Craig's study (1994). Young et al. (2012) administered the Rorschach Test to 15 child sexual abusers and 45 rapists and found that rapists demonstrated more disordered thinking and had less interest in forming psychological attachments with others. In comparison, the child sexual abusers demonstrated more logical thinking and had a greater need for emotional attachment, a more immature self-focus, and greater feelings of emotional alienation from others.

Compared to the empirical research conducted on child sexual abusers and rapists, research on internet sexual offenders is sparse despite the notable increase in convictions

of this type of sex offense (Laulik et al., 2007). Internet sexual offenders can be classified by the specific type of offense they commit, including the viewing of child pornography, distributing child pornography, or contacting children via the internet to establish opportunities to sexually offend (Seto et al., 2011). Regarding demographic characteristics, internet sexual offenders are predominantly male and of European descent (Webb et al., 2007). However, other demographic variables such as age and level of income are mixed across this group of offenders, although they tend to be younger than child sexual abusers (Reijnen et al., 2009; Webb et al., 2007). Research has also found that internet sexual offenders live alone more often and have fewer live-in relationships than child sexual abusers (Reijnen et al., 2009; Webb et al., 2007). When comparing the developmental histories of child sexual abusers to internet sexual offenders, Webb et al. (2007) found that both groups experienced substantial amounts of childhood difficulties; however, the child sexual abuser sample reported significantly more physical abuse in childhood.

To date, studies have not evaluated psychiatric diagnoses that are most common among internet sexual offenders and the prevalence rate of pedophilia among this group is currently unknown (Simons, 2015). However, numerous studies have assessed the personality functioning and characteristics of this group by utilizing various personality measures, similar to child sexual abusers and rapists. Laulik et al. (2007) administered the Personality Assessment Inventory (PAI) to a sample of 30 convicted internet sexual offenders and found impairment in interpersonal functioning and difficulties with emotional regulation. Specifically, a significant portion of the sample appeared underassertive, lacking in empathy, and self-conscious in interpersonal interactions. A

study by Magaletta et al. (2012) largely confirms these results, finding that their sample of internet sexual offenders, specifically child pornography offenders (n = 35), had affect regulation difficulties and interpersonal deficits. Other findings in this study suggest that internet sexual offenders presented with low levels of aggression and hostility and high levels of stress. Using the Minnesota Multiphasic Personality Inventory-Second Edition (MMPI-2), Tomak et al. (2009) found that internet sexual offenders (n = 48) differ from contact sexual offenders (n = 104) in that they present as less physically aggressive, less deviant, and less impulsive.

Although sex offender typologies seek to identify homogenous subgroups of offenders based on offense type, rapist, child sexual abuser, and internet sexual offender populations remain as relatively heterogenous groups (Magaletta et al., 2012; Hall et al., 1986; Craissati, 2005). Despite this, there appears to be some personality characteristics that are prominent across all subgroups of sex offenders. Although sex offenders commit sex offenses for a multitude of reasons, engagement in these sexual acts indicate that these offenders have failed to achieve intimacy and are unable to meet their emotional and sexual needs in a pro-social manner (Beech et al., 2008). Consistent research findings support that all typologies of sex offenders frequently experience a broad range of serious social deficits and interpersonal problems. Further, this population often lacks social skills that are necessary in order to develop and maintain adequate relationships with others. Studies have also found that sex offenders often experience high levels of loneliness and social isolation when compared to non-offenders, likely due to their difficulties in establishing relationships and engaging with others (Maniglio, 2012).

Evaluation of Sex Offenders

There is a wide variety of clinical and legal contexts in which sex offenders undergo evaluations. Depending on the referral question, sex offenders may receive an evaluation to assess their level of dangerousness, risk for recidivism, need for involuntary commitment, amenability to treatment, or to determine their supervision and treatment related needs. These evaluations can take place at various stages in criminal justice proceedings including preadjudication, at sentencing, prior to release from incarceration, or when receiving treatment in the community. Alleged sex offenders are frequently referred to receive an evaluation by their criminal defense attorney to aid in evaluating the offender's risk to the community and to develop a treatment or management plan (Rogers & Bender, 2018). The outcome of these evaluations can be used to determine criminal sentencing. Due to the high-stakes decisions that can result from these evaluations, evaluators are called on to perform a comprehensive and individualized evaluation that utilizes multiple data sources and empirically derived instruments for assessing sex offenders (McGrath & Purdy, 1999). The evaluation process typically includes a clinical interview, review of records (e.g., police reports, prior treatment records, medical records, court documents), and a battery of tests that typically assess for level of risk, psychopathology, and deviant sexual interest and behaviors. Evaluators typically administer a combination of instruments that often include sex offender specific risk assessments, personality assessments, other psychometric instruments, and physiological measures, depending on the referral question (Drogin et al., 2011).

A critical question that can arise during a sex offender evaluation in the preadjudication stage is whether or not the alleged sex offender actually committed the

sexual offense they have been charged with (McGrath & Purdy, 1999). Many evaluees who undergo these evaluations deny their engagement in the sexual offense for which they are being referred. It is not uncommon for attorneys or other criminal justice professionals to refer alleged offenders to complete an evaluation in the process of determining their guilt or innocence. However, the practice guidelines produced by the Association for the Treatment of Sexual Abusers (ATSA; 2005) state, "evaluators do not offer conclusions regarding whether an individual has or has not committed a specific act of sexual abuse" (p. 11). Therefore, it is not the evaluator's duty to be the finder of facts for the case and it is unethical for him or her to offer an opinion regarding the evaluee's guilt or innocence (Drogin et al., 2011).

Historically, evaluating the level of risk a sex offender poses to the community was solely based on information that was gathered from an unstructured clinical interview (Tarescavage et al., 2018). The clinical interview is used to collect information from the evaluee regarding family of origin, criminal justice involvement, medical history, romantic relationships, and sexual interests and behaviors (Phenix & Hoberman, 2016). This approach has since been criticized as having questionable reliability and validity as the evaluee's self-report is often biased. Research has found that sex offenders undergoing evaluations frequently minimize or deny their sexual interests, behaviors, and general psychological symptoms in order to portray themselves as well-adjusted. The interview is also limited by the influence of the evaluator's own biases (Tarescavage et al., 2018). Due to such limitations of an unstructured clinical interview approach, more objective, empirically derived measures have been created to assist in the evaluation process. As the primary focus of many sex offender evaluations are to predict an

offender's likelihood of sexual offense recidivism, numerous actuarial measures have been developed to measure risk factors for re-offending. Sex offender specific actuarial risk assessment measures were created by identifying variables associated with sexual re-offending, which were common among known sex offender populations (McGrath & Purdy, 1999). These assessments typically evaluate static risk factors, dynamic risk factors, or a combination of both. Static risk factors are features of the sex offender's history that predict recidivism and are not amenable to clinical intervention; they include age, gender, history of prior non-sexual offenses, and history of prior sexual offenses. Dynamic risk factors, such as employment status, active substance abuse, and peer group influences, are potentially changeable factors and are amenable to intervention efforts (Baldwin, 2015). Hansen and colleagues (Hansen & Bussiere, 1998; Hanson & Morton-Bourgon, 2005) conducted meta-analyses on the utility of various risk assessment instruments and found that the Static-99, Static-2002, MnSOST-R, Risk Matrix 2000-Sex, and SVR-20 were the best supported measures of assessing the likelihood of sexual offense recidivism (Baldwin, 2015). The Static-99 and its revised editions are the most widely used and researched sex offender risk assessments in the world (Harris & Hanson, 2010). The Static-99R is a 10-item actuarial assessment developed to evaluate adult male sex offenders' risk for sexual offense recidivism. The Static-99R is only suitable for use if an individual has been charged or convicted of at least one sex offense against a child or non-consenting adult. The 10 items assess for prior sex offenses, prior sentencing, convictions for non-contact sex offenses, index non-sexual violence convictions, prior non-sexual violence convictions, unrelated victims, stranger victims, male victims, lack of a long-term intimate relationship, and if the offender is 25-years-old or under upon

release (Beech et al., 2009). The total score derived from the summation of the 10 items will provide the evaluator with one of five possible risk categories that the evaluee meets – Very Low Risk, Below Average Risk, Average Risk, Above Average Risk, and Well Above Average Risk. Because the Static-99R and other actuarial instruments are based on empirically validated risk indicators, scoring is objective and the evaluator follows specifically stated rules. These measures are less influenced by evaluator and evaluee bias because risk assessment items can typically be answered by reviewing judicial records (Phenix & Hoberman, 2016). Although risk assessments cannot predict with absolute certainty that a sex offender will or will not re-offend, actuarial risk assessments have demonstrated the ability to accurately classify recidivists at levels well above chance (McGrath & Purdy, 1999).

A number of psychometric tests have been developed and are commonly used in sex offender evaluations (Beech et al., 2009). These measures typically assess the sex offender's attitudes, beliefs, and values and also survey their sexual interests and behaviors. The data collected from these tests can assist evaluators in determining the offender's treatment needs and potential risks, and provide diagnostic clarification. Some of the most commonly used psychometric tests to assess for deviant sexual interests include the Multiphasic Sex Inventory (MSI; Nichols & Molinder, 1984) and the Derogatis Sexual Functioning Inventory (DSFI; Derogatis, 1978). The MSI and its revised edition, the MSI-II (Nichols & Molinder, 2000), are the most widely used self-report questionnaires designed to assess sex offenders' psychosexual characteristics, specifically their sexual activities, problems, and experiences (Phenix & Hoberman, 2016). These measures provide information that is independent from psychopathology

and personality assessments. The original MSI designed to be used for adult male sex offenders contains 20 scales – the core scales consist of the validity and paraphilia scales. The validity scales identify whether the sex offender is responding in a socially desirable way, exaggerating his or her problems, or denying his or her problems. The paraphilia scales specifically measure the sex offender's deviant sexual interest and offending behaviors such as pedophilic interests, voyeurism, exhibitionism, and sado-masochism. The MSI's most recent revised edition, the MSI-II, expanded to include separate measures for adult female sex offenders, adult male sex offenders, adolescent female sex offenders, and adolescent male sex offenders. Additionally, the revised edition of this measure can now be utilized with alleged sex offenders who either admit to or deny their sexual assault or misconduct allegations. The MSI-II Adult Male Form consists of 560 true/false questions and includes 12 separate measures of test reliability and validity that assess for the sex offender's test taking approach and response patterns. New scales were developed for this revised addition to assess for a broader range of paraphilias, to examine an offender's justification for his or her sexual offense, and to examine for potential gender identity problems (Phenix & Hoberman, 2016). Unique to the MSI-II is that the evaluee's scores can be compared to scores of known child molesters, rapists, and individuals with specific paraphilias. These psychometric self-report measures allow for a more comprehensive assessment of sex offenders' sexual histories and deviant attitudes and behaviors, in addition to information retrieved during the clinical interview. Zonana et al. (2004) suggested that sex offenders are more likely to reveal their dysfunctional sexual behaviors or perceptions on these measures than they would provide during a face-to-face interview.

Physiological assessments are also commonly used in sex offender evaluations. Sex offenders are often motivated to minimize or deny their deviant sexual interests as admitting to these interests may result in harsher sentences or restrictions imposed on them. Therefore, evaluators incorporate objective measures of deviant sexual interests into their evaluations (Coric et al., 2005). Research has shown that the existence of deviant sexual interests and arousal are strong predictors of sexual re-offending (Beech et al., 2009). Thus, it is imperative that evaluators acquire accurate information related to offenders' sexual thoughts, fantasies, and urges.

The three most commonly used methods of physiological assessment used to evaluate sex offenders include the penile plethysmography (PPG), polygraph test, and viewing time measures (Drogin et al., 2011). The PPG is one of the oldest and most controversial instruments that measure sex offenders' sexual interests. The PPG is an objective measure of a male's physical sexual arousal when he views or listens to deviant sexual material, non-deviant sexual material, or neutral material (Phenix & Hoberman, 2016). The visual or audio stimuli presented to the evaluee contains individuals of various ages and genders and various sexual and nonsexual encounters. The two most common types of PPGs used today measure changes in either the penile volume or circumference (Coric et al., 2005). The stimuli resulting in the greater penile tumescence likely reflects the offenders' sexual preferences or interests. Since its development, the PPG has received criticism for its lack of standardized testing and scoring procedures, thus threatening the validity and reliability of the test results. Additionally, researchers have found that sex offenders have been able to suppress their sexual arousal by looking away from the presented visual stimulus or engaging in a non-sexual activity to distract

themselves, compromising the results of the test (Rogers & Bender, 2018). Evaluators have also argued that the PPG is intrusive and potentially an unethical method of assessing sexual deviancy. Currently, courts in the United States have not reached a uniform agreement as to the admissibility of the PPG (Murphy et al., 2020).

The polygraph test is another physiological approach that is widely used to assess sex offenders in the United States (Rogers & Bender, 2018). In a typical polygraph examination, the evaluee is read questions posed by the evaluator while the polygraph device records the changes in the evaluee's physiological response (e.g., blood pressure, pulse, respiration, skin conductivity). Due to the limitations of this measure's validity and reliability, polygraph results are not admissible in U.S. Courts (Phenix & Hoberman, 2016). The polygraph test is more commonly used for post-conviction evaluations or to assess adjudicated sex offenders' treatment compliance and progress. This measure is also given to evaluate the accuracy of information that the sex offender disclosed regarding sexual background, particularly his or her history of sexually deviant behavior (Beech et al., 2009). The results of the polygraph test can assist in obtaining a more comprehensive history of his or her sexual offenses and potentially aid in increasing the offender's number of disclosures. Research has found that polygraph examinations lead to considerably more disclosures (i.e., in the number of victims perpetrated against, the range of offending behavior) (Phenix & Hoberman, 2016). However, ATSA (2014) advises evaluators not to use polygraph test results as the only source of data when conducting an evaluation, but rather in conjunction with other assessment findings.

Viewing time measures are one of the newer and less invasive approaches designed to assess for deviant sexual interests and attraction of sex offenders (Phenix &

Hoberman, 2016). Viewing time is a measure of the length of time an evaluee takes to view images of various individuals; the underlying theory is that the longer an evaluee takes to look at the image, the more he or she is sexually aroused or interested in the material. The most widely used viewing time analysis measure is the Abel Assessment of Sexual Interest-Second Edition (AASI-2; Abel, 1995). The AASI contains a self-report questionnaire that obtains information on the evaluee's sexual history, preferences and behaviors, and legal history. Another component of the AASI requires the evaluee to view 160 slides displaying individuals of various ages, races, and genders. The slides do not contain nudity and are not overtly sexual in nature; however, the slides were designed to be provocative to the evaluee who has sexual interests in one of the targeted categories (Sachsenmaier & Grees, 2009). Visual reaction time is recorded without the evaluee knowing when he or she is viewing the images. The evaluee is asked to rate each image based on the level of sexual attractiveness he or she experiences using a scale ranging from 1 to 7. A score of 1 indicates a very low interest and high disgust related to the image while a score of 7 indicates that the image is very sexually arousing to them (Phenix & Hoberman, 2016). Similar to the PPG and polygraph test, the validity and reliability of the results produced from the AASI have largely been debated in the literature (Coric et al., 2005). Due to these reasons, AASI results are not admissible in all U.S. Courts (Rogers & Bender, 2018).

Personality Testing in Sex Offender Cases

In addition to the previously mentioned measures, personality assessments are frequently given in sex offender evaluations. In order to understand the purpose and information gained from a personality assessment, it is necessary to understand what

these instruments attempt to measure. Personality is commonly defined as "individual differences in characteristic patterns of thinking, feeling, and behaving" (American Psychological Association, 2000, para. 1). Cohen and Swerdlik (2009) expanded upon this definition, describing personality as "an individual's unique constellation of psychological traits and states" that include attitudes, interests, values, personal identity, and behavioral and cognitive styles (p. 391). Personality traits are consistent and stable patterns of behavior that differ in strength among people. Individuals can be distinguished from one another not based primarily on whether they have certain traits, but by the intensity or pattern of traits they have (Harwood et al., 2011). In contrast, personality states are less stable over time and can change in response to an individual's environment or circumstances. Thus, traits are enduring characteristics of behavior that are internally caused and transcend changes in the social environment, whereas states are situation-specific and externally caused (Harwood et al., 2011).

Personality assessments can assist in identifying an individual's unique constellation of traits and states. Personality assessments are used to evaluate these personality characteristics to guide decision making in clinical, forensic, health care, educational, and organizational settings (Weiner & Greene, 2008). These measures were traditionally used in clinical settings to assist in differential diagnoses and to inform treatment planning. Harwood et al. (2011) discussed six clinically relevant domains that personality assessments can provide information on including:

> (1) The individual's diagnosis or disorder; (2) the etiology or causes of the disordered behavior; (3) the prognosis or anticipated course of this problem; (4) the nature of the treatments that may ameliorate or alter that course or prognosis;

> (5) the degree of functional impairment in both routine and specialized life functions; and (6) the person's pattern of strengths and adaptive capacities (p. 3).

Many of the popular personality assessments designed for clinical use are also frequently used in forensic evaluations (Archer, 2006). However, these instruments typically require adjustments to their administrative and interpretive procedures when working with this specialized population. Despite the large body of tests that have been created specifically for forensic usage, survey findings completed by clinical psychologists consistently show that traditional clinical assessments are the most frequently used measures in forensic evaluations (Archer, 2006). The three commonly used approaches of personality assessment are self-report questionnaires, performance-based methods, and behavioral assessments, which will be discussed in a later section.

Personality assessments are utilized to inform forensically-relevant decisions in various types of forensic evaluations. Types of forensic evaluations include but are not limited to: competency to stand trial, criminal responsibility, personal injury, child custody, civil commitment, dangerousness, and sex offender evaluations (Archer et al., 2016). Researchers have surveyed psychologists and psychiatrists who conduct these evaluations to identify the most commonly used assessments. Borum and Grisso (1995) surveyed experienced forensic psychologists and psychiatrists regarding their usage of psychological tests in evaluations for criminal responsibility and competency to stand trial. Regarding criminal responsibility evaluations, forensic evaluators most frequently used the MMPI or MMPI-2, a self-report personality measure; these tests are commonly used by 94% of the survey respondents. Forty-two percent of these respondents reported using performance-based methods such as the Rorschach Test when evaluating for

criminal responsibility. Similar results were found when evaluators described their most commonly used measures when conducting competency to stand trial evaluations; 90% of evaluators commonly used self-report measures, while 33% commonly used performance-based measures. A more recent study conducted by Archer et al. (2006) surveyed 152 clinical psychologists who practice in the realm of forensic psychology and have conducted forensic evaluations. The results of the study found that the most commonly used self-report questionnaire used in forensic evaluations with adults was the MMPI-2, followed by the Personality Assessment Inventory (PAI; Morey, 1991), and the Millon Clinical Multiaxial Inventory – Third Edition (MCMI-III; Millon, 1994), respectively. The Rorschach Test was found to be the most frequently used performance-based measure of personality; however, this test was not as frequently used compared to the self-report measures listed. When compared to single-scale tests, cognitive/intellectual tests, neuropsychological tests, and other specialized forensic instruments, self-report measures were used by the largest portion of respondents, with 86% of respondents having reported the use of one or more of these instruments. Performance-based measures were used the least among the various measures, with 36% of respondents reporting usage (Archer et al., 2016).

Traditional personality assessment instruments are beneficial to use in forensic evaluations as they typically assess for an evaluee's aberrant response style (i.e., under-reporting, over-reporting) and the evaluee's functioning in numerous domains. Additionally, these psychological tests often have better established validity and reliability due to the extensive research that has been conducted on them over the years, compared to the more recently developed specialized-forensic instruments (Archer et al.,

2016). However, because these personality assessments were not specifically designed for forensic use, few of them have appropriate forensic norms and the results of the measure cannot be used to answer specific legal questions that the evaluator may have been asked to answer (Archer, 2006). The American Psychological Association's Specialty Guidelines for Forensic Psychology state that:

> Assessment in forensic contexts differs from assessment in therapeutic contexts in important ways that forensic practitioners strive to take into account when conducting forensic examinations. Forensic practitioners seek to consider the strengths and limitations of employing traditional assessment procedures in forensic examinations (American Psychological Association, 2013, p. 15).

Archer (2006) recommends evaluators utilize clinical assessment measures only when they are appropriate and relevant to the legal question and that these measures should never be given in isolation, suggesting additional sources of information be collected to better answer the referral question.

Specifically regarding sex offender evaluations, personality tests are frequently given to measure distinct personality features and the degree of psychopathology present in the evaluee. These tests are not designed to specifically evaluate for past, current, or future deviant sexual interests and behaviors; rather, these tests assess for more general personality dysfunction (Marshall & Hall, 1995). The results of the personality measures can aid in dispositional decisions by providing the diagnosis or diagnoses (if applicable) of the sex offender and treatment implications. Consistent with the findings of national surveys regarding the forensic applications of personality measures, the MMPI and its revised editions are the most commonly used self-report measures with the sex offender

population and will be discussed at length in future sections. Following the MMPI, the PAI and MCMI are the other most commonly used self-report questionnaires to evaluate sex offenders (Davis & Archer, 2010). The Rorschach Test is among the most commonly used performance-based measures utilized in forensic settings, including sex offender evaluations (Archer, 2006).

The Personality Assessment Inventory (PAI) was published in 1991 as a multidimensional, self-report inventory designed to assess for abnormal personality traits in adults. The PAI consists of 344 items scored on a 4-point Likert scale organized under 22 non-overlapping scales - four validity scales, 11 clinical scales, five treatment scales, and two interpersonal scales. The validity scales evaluate the evaluee's response style on the test and can identify if the evaluee responds in an inconsistent or random manner, and if he or she attempts to exude a positive (i.e., fake-good) or negative (i.e., fake-bad) impression. The 11 clinical scales provide information on a wide range of clinical constructs that can be divided into three broad classes of disorders that include the neurotic cluster, psychotic cluster, and behavioral disorders cluster. Ten of these clinical scales contain subscales that are conceptually derived. The five treatment scales assess for personality traits or behaviors that could serve as barriers to treatment such as aggressive tendencies, suicidal ideation, and lack of social support. The two interpersonal scales assess for an individual's style of social engagement on two domains – warm affiliation versus cool rejection and dominance versus submissiveness (Morey, 1991).

The PAI is commonly used in forensic evaluations for a number of reasons. First, the PAI is a relatively short measure compared to other personality measures such as the MMPI and requires a relatively low reading level (4^{th} grade). This increases the

accessibility of the test as many offenders involved in the criminal justice system have lower levels of educational achievement (Edens et al., 2001). Additionally, this measure is psychometrically sound, extensively researched, and faces few admissibility challenges in court (Archer, 2006; Thorpe & Dawson, 2010).

The Millon Clinical Multiaxial Inventory (MCMI) was published in 1977 and has undergone four revisions. The MCMI-IV is the most recently published edition and was intended to assess personality disorders and prominent psychiatric syndromes (Archer et al., 2016). While many of the clinical scales on this self-report measure are consistent with personality disorders from the Diagnostic and Statistical Manual – Fifth Edition (DSM-5; American Psychiatric Association, 2013) in both name and content, other scales are rooted in Millon's evolutionary theory of personality psychopathology (Rouse, 2017). The MCMI-IV consists of 195 true/false questions organized under 30 scales – five validity scales, 15 personality scales, and 10 clinical syndrome scales. Similar to the PAI, the MCMI-IV contains validity scales that evaluate for random or inconsistent responding and attempts by the evaluee to underreport or overreport his or her symptoms. The 15 personality scales are composed of 12 basic personality patterns and three severe personality pattern scales. The basic clinical personality pattern scales assess for core characteristics of the evaluee's functioning that are pervasive and stable. The severe personality pattern scales assess for acute and advanced stages of personality pathology. Unlike the basic personality scales, these scales reflect a marked deterioration in personality structure that include deficiencies in social competence and periodic psychotic episodes (Grossman & Amendolace, 2017). Forty-five facet scales were developed to correspond with the personality scales in order to provide a more specific

presentation of the evaluee's functioning. Similar to the personality scales, the clinical syndrome scales are separated into two groups, seven scales assessing for moderate levels of psychopathology, and three scales assessing for severe levels of psychopathology. The clinical syndrome scales aid in identifying the evaluee's temporary and transient states compared to the more stable personality pattern scales (Zachar, 2017).

The MCMI and following revisions have a number of strengths that make this test popular among forensic evaluators. The MCMI test scales demonstrate reliability and internal consistency, while the test manual provides clear scoring rules for obtaining scale scores (Archer et al., 2016). Despite these strengths, the MCMI has remained controversial in forensic settings. One of the predominant criticisms of the MCMI is that the test was standardized solely on a clinical sample and thus, does not allow for discrimination between clinical and normal populations. Moreover, this measure can potentially overpathologize evaluees undergoing forensic evaluations who do not have clinical symptomatology present (Archer, 2006).

The Rorschach Test is a performance-based measure that was published in 1921 and assesses for the dynamic and structural aspects of personality functioning (Weiner, 2003). The Rorschach Test is comprised of 10 standard inkblot plates that are presented to the evaluee one at a time. How the evaluee responds to each stimulus will provide information on his or her capacities and personality characteristics. In recent decades, Exner's Rorschach Comprehensive System has been the most commonly used standardized method of administering, scoring, and interpreting the Rorschach Test, supplemented more recently with the Rorschach Performance Assessment System (R-PAS). The results of the Rorschach Test can assist in identifying the evaluee's thought

processes, coping abilities, affective expressions, self-perception, and interpersonal capacities (Krishnamurthy et al., 2011). There are many advantages to using this performance-based measure in forensic evaluations. Compared to self-report inventories, the Rorschach Test is relatively unstructured and ambiguous. Therefore, this measure is more resistant to an evaluee's impression management strategy (i.e., attempts to over-report or under-report symptoms) as he or she has little direction on what the test measures and how his or her responses are scored (Archer et al., 2016). Additionally, the Rorschach Test is relatively culture-free and is a psychometrically sound measure when administered and interpreted properly. McCann (1998) asserted that the Rorschach Comprehensive System meets the U.S. Court's legal and professional standards of test admissibility due to this measure's standardized administration and scoring procedures, psychometric characteristics, and large body of literature in support of this measure. Weiner et al. (1996) found that the Rorschach Test faces little admissibility challenges in court.

Overview of the MMPI, MMPI-2, MMPI-2-RF

The MMPI (Hathaway & McKinley, 1943) and its revised editions, including the MMPI-2 (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001), and MMPI-2-RF (Ben-Porath & Tellegen, 2008), are the most widely used and researched self-report personality questionnaires worldwide. The widespread use of these personality measures can be attributed to several factors including their simplicity of administration and scoring, large item pool, inclusion of validity scales to assess response styles, strong empirical support, and application of use in a variety of settings.

The MMPI was published in 1943 to aid in assessing psychopathology and personality characteristics in adults. The original self-report measure consisted of 566 items answered in a true or false format, yielding three validity scales and ten standard clinical scales (Friedman et al., 2015). The validity scales were developed to assist the evaluator in detecting deviant test-taking attitudes (i.e., "faking-good," "faking-bad") that reduce or invalidate the interpretability of the test findings (Greene, 1980). The three validity scales consisted of the Lie scale (L), Infrequency scale (F), and Defensiveness scale (K). The L scale consisted of 15 items that were designed to detect an unsophisticated attempt by the respondent to present himself or herself in an unrealistically positive light. The F scale consisted of 64 items that were developed to detect exaggeration or overreporting of psychological disturbance and distress. The K scale consisted of 30 items designed to measure a respondent's denial of psychopathology. In addition to the standard validity scales, other validity measures were developed for the MMPI to detect underreporting of symptoms including the Positive Malingering (Mp) scale (Cofer et al., 1949) and the Wiggin's Social Desirability (Wsd) scale (Wiggins, 1959). The Positive Malingering (Mp) scale consisted of 33 items that measure one's denial of common flaws and exaggeration of positive adjustment, similar in function to the Lie (L) scale. The Wiggin's Social Desirability (Wsd) scale consisted of 40 items that measure one's over endorsement of positive attributes (Friedman et al., 2015).

Ten standard clinical scales were developed to assist the evaluator in determining the presence, type, and severity of psychopathology conveyed by the respondent through his or her responses to test items. These clinical scales included Scale 1, Hypochondriasis

(Hs), a measure of bodily complaints or somatic concerns; Scale 2, Depression (D), a measure of symptomatic depression; Scale 3, Hysteria (Hy), a measure of denial and repression of physical, psychological, and emotional difficulties; Scale 4, Psychopathic Deviate (Pd), a measure of antisocial inclinations and social alienation; Scale 5, Masculinity-Femininity (Mf), a measure of the adherence to or departure from stereotypic gender roles; Scale 6, Paranoia (Pa), a measure of interpersonal sensitivity and suspiciousness of others; Scale 7, Psychasthenia (Pt), a measure of psychological turmoil and discomfort; Scale 8, Schizophrenia (Sc), a measure of the presence of bizarre thought processes and other common symptoms of schizophrenia; Scale 9, Hypomania (Ma), a measure of mood elevation and psychomotor excitement commonly associated with hypomanic or manic episodes; and Scale 0, Social Introversion (Si), a measure evaluating the respondent's level of interpersonal comfort and desire to engage in social exchanges (Friedman et al., 2015).

In addition to the validity scales and standard clinical scales, a multitude of subscales, content scales, and supplementary scales were developed by researchers to further aid evaluators in the interpretation of test results. As the MMPI item pool was heterogenous in nature, the ten clinical scales were diverse in content (Friedman et al., 2015). The Harris-Lingoes subscales (Harris & Lingoes, 1955; 1968) were developed for clinical scales 2, 3, 4, 6, 8, and 9 to create more homogenous groups of test items to increase the interpretive depth of the clinical scales. These subscales were created by inspecting the item content for each clinical scale and grouping items into content categories that reflected a similar trait or attribute. Harris and Lingoes did not develop subscales for clinical scales 1 and 7 as the items associated with these scales were

homogenous in content. Although scales 5 and 0 were multidimensional, they did not receive Harris-Lingoes subscales because they were viewed as nonclinical scales. Additionally, Wiggins (1966) designed 13 content scales using intuitive and psychometrically sound procedures to group items together that share similar item content. Wiggins designed each content scale to be a homogenous measure of a particular dimension that did not overlap with other established scales. These homogenous face-valid scales reflected the respondent's self-report of symptomatology (Greene, 1980).

Since the development of the MMPI, a wide range of research-based supplementary scales were developed to further assess for more specific facets of psychopathology. However, of the numerous supplementary scales that were created, only eight of these scales were incorporated into the standard test scales on the MMPI. In addition to the MMPI scales, lists of empirically-derived critical items were created. These critical items were identified to signal severe psychopathology and imminent risks of harm that may require immediate clinical attention. Koss and Butcher (1973) and Lachar and Wrobel (1979) are the most notable developers of these critical items and these item lists have continued to be used on later revisions of the MMPI.

The MMPI-2 (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) is the first revision of the original MMPI and is currently in use as the MMPI-3 was released only recently in November, 2020. The re-standardization of this test was needed in order to provide updated test norms, develop a larger normative sample, revise item content, and include new content areas. To maintain continuity between the MMPI and MMPI-2, test developers made minimal revisions to the original validity scales and

clinical scales except for the rewording of 68 items and the elimination of 13 items due to objectionable content (Greene, 2000).

The MMPI-2 consists of 567 true/false items and retained the previous three validity scales and ten standard clinical scales that were included in the original MMPI. However, three additional validity scales were added to this revised edition including the Variable Response Inconsistency scale (VRIN), the True Response Inconsistency scale (TRIN), and the F-Back scale (FB). The VRIN scale is comprised of 67 item pairs designed to detect contradictory and inconsistent responding while the TRIN scale consists of 23 item pairs designed to detect biased responding in the acquiescent (yea-saying) or non-acquiescent (nay-saying) direction. The FB scale contains 40 F scale items placed in the latter half of the test, designed to detect random or biased responding that predominantly affects the supplementary and content scale scores. Three additional validity scales including the Infrequency-Psychopathology scale F(p), Fake Bad Scale (FBS), and the Superlative Self-Presentation (S) scale were added to the MMPI-2 profile form after this revision was published. The F(p) scale contains 27 items designed to be used conjointly with the F scale to detect overreporting of psychopathology. Previously named the Fake Bad Scale (FBS), the Symptom Validity Scale consists of 43 items that also assist in the detection of overreporting psychological problems. The Superlative Self-Presentation (S) scale consists of 50 items and was developed to measure a respondent's level of defensiveness and tendency to present an excessively positive self-portrayal. The Positive Malingering (Mp) scale and Social Desirability (Sd) scale, which were supplementary validity measures for the MMPI, were revised to be similarly available for the MMPI-2. The 10 standard clinical scales developed for the MMPI

remained virtually the same on the revised edition, the difference being that a few items were removed from scales 1, 2, 5, and 0 due to objectionable item content and other items were revised to improve clarity of wording (Friedman et al., 2015).

In addition to the validity scales and standard clinical scales, the MMPI-2 consists of the original set of 28 Harris-Lingoes subscales, 15 new content scales, 15 supplementary scales – eight of which were retained from the MMPI, new Personality Psychopathology Five (PSY-5) scales, and nine new restructured clinical (RC) scales. New content scales were created to assess the four major content themes in the MMPI-2 including internal symptoms, external or aggressive tendencies, a devalued view of self, and general problem areas. The supplementary scales include a subset of previously developed scales, in addition to newly developed ones that assess for personality characteristics, emotional distress, behavioral dyscontrol, and gender roles. The PSY-5 scales were developed to assess for five domains of disordered personality including Aggressiveness (AGGR), Psychoticism (PSYC), Disconstraint (DISC), Negative Emotionality/ Neuroticism (NEGE), and Introversion/Low Positive Emotionality (INTR). The restructured clinical (RC) scales were created and later added to the MMPI-2, assisting in the examination of the substantive core of each clinical scale after extracting the general maladjustment or distress domains commonly shared by the scales (Friedman et al., 2015).

The MMPI-2-RF (Ben-Porath & Tellegen, 2008) is the most recent revision of the MMPI and was created to be an alternative measure, but not a replacement, for the MMPI-2. After the development and implementation of the restructured clinical (RC) scales on the MMPI-2, test developers concluded that a restructuring of the entire test

should be undertaken. The MMPI-2-RF contains 338 true/false items, all of which were taken from the 567-item MMPI-2. Thus, no new items were added on this measure. This self-report measure consists of nine RC scales previously included on the MMPI-2, nine validity scales, three higher order scales, 23 specific problems scales, two interest scales, and the revised PSY-5 scales.

The MMPI-2-RF validity scales consist of the revised versions of the eight scales previously used on the MMPI-2, and one new validity scale specifically created for this revised edition to detect overreporting of difficulties. These nine validity scales assist the evaluator in evaluating for inconsistent responding (VRIN-r, TRIN-r), overreporting of psychological problems (F-r, Fp-r, Fs, FBS-r, RBS), and underreporting of psychological problems (L-r, K-r). Table 1 presents a comparison of the validity scales for each version of the MMPI.

Table 1

Comparisons of the MMPI, MMPI-2, and MMPI-2-RF validity scales

MMPI	MMPI-2	MMPI-2-RF
Cannot Say score (?)	Cannot Say score (?)	Cannot Say score (?)
Infrequency scale (F)*	Variable Response Inconsistency scale (VRIN)	Variable Response Inconsistency scale (VRIN-r)
Lie scale (L)*	True Response Inconsistency scale (TRIN)	True Response Inconsistency scale (TRIN-r)
Defensiveness scale (K)*	Infrequency scale (F)*	Infrequent Responses (F-r)*
	Back Infrequency scale (FB)	Infrequent Psychopathology Responses (Fp-r)
	Infrequency-Psychopathology scale (Fp)	Infrequent Somatic Responses (Fs-r)
	Lie scale (L)*	Symptom Validity (FBS-r)
	Correction scale (K)*	Uncommon Virtues (L-r)*
	Superlative Self-Presentation scale (S)	Adjustment Validity (K-r)*

Note. * denotes scales that have been retained across all test revisions with modifications.

Three higher-order scales were developed for the MMPI-2-RF to provide a dimensional assessment of psychopathology; they are the Emotional/Internalizing Dysfunction (EID) scale, the Thought Dysfunction (THD) scale, and the Behavior/Externalizing Dysfunction (BXD) scale. The EID scale contains 41 items that measure general subjective distress and negative affect (Greene, 2000). The THD scale consists of 26 items that measure thought dysfunction. The BXD scale contains 23 items that measure behavioral acting-out tendencies.

The nine RC scales that constitute the core set of scales on the MMPI-2-RF were retained and are identical to the RC scales on the MMPI-2. The RC scales do not replace the clinical scales of the MMPI and MMPI-2. Rather, they measure the core constructs of these clinical scales through elimination of their shared variance (Friedman et al., 2015). This was done by removing the overlapping items between the clinical scales to increase the specificity of the measure. These RC scales include RCd, Demoralization, a measure of general distress and emotional discomfort; RC1, Somatic Complaints, a measure of preoccupation with and complaints of physical functioning; RC2, Low Positive Emotions, a measure of negative emotionality commonly associated with depression and anxiety; RC3, Cynicism, a measure of mistrust of others; RC4, Antisocial Behavior, a measure of impulsivity and acting-out behaviors; RC6, Ideas of Persecution, a measure of thought problems and suspicious thinking; RC7, Dysfunctional Negative Emotions, a measure of maladaptive emotional experiences; RC8, Aberrant Experiences, a measure of unusual perceptions or thoughts; and RC9, Hypomanic Activation, a measure of over-activation. RC scales correspond with clinical scales 1, 2, 3, 4, 6, 7, 8, and 9. RC scales were not created for clinical scales 5 and 0 as they are not deemed to reflect

psychopathology (Friedman et al., 2015). The RCd scale does not correspond with a clinical scale as it was created to measure general distress that was previously found across all clinical scales. Table 2 depicts the clinical scales across the different MMPI versions.

Table 2

The standard clinical scales of the MMPI/MMPI-2 and the corresponding restructured scales of the MMPI-2/MMPI-2-RF

MMPI/MMPI-2	MMPI-2/MMPI-2-RF
-----	RCd, Demoralization
Scale 1, Hypochondriasis (Hs)	RC1, Somatic Complaints
Scale 2, Depression (D)	RC2, Low Positive Emotions
Scale 3, Hysteria (Hy)	RC3, Cynicism
Scale 4, Psychopathic Deviate (Pd)	RC4, Antisocial Behavior
Scale 5, Masculinity-Femininity (Mf)	-----
Scale 6, Paranoia (Pa)	RC6, Ideas of Persecution
Scale 7, Psychasthenia (Pt)	RC7, Dysfunctional Negative Emotions
Scale 8, Schizophrenia (Sc)	RC8, Aberrant Experiences
Scale 9, Hypomania (Ma)	RC9, Hypomanic Activation
Scale 0, Social Introversion (Si)	-----

The Specific Problems (SP) scales were created for the MMPI-2-RF to identify important characteristics or clinically-relevant issues that are subsumed under or associated with particular elevations of RC scales. The SP scales are organized into four subsets: Somatic/Cognitive scales, Internalizing scales, Externalizing scales, and Interpersonal scales. The Somatic/Cognitive scales consist of five individual scales designed to evaluate preoccupation with one's physical health, somatic problems, and cognitive complaints – Malaise, Gastrointestinal Complaints, Head Pain Complaints, Neurological Complaints, and Cognitive Complaints. The Internalizing scale contains

nine individual scales that measure a wide range of emotional dysfunction. This scale assists in identifying more specific areas of interest related to mood and affect problems subsumed under the EID Higher-Order scale in addition to the RCd, RC2, and RC7 scales – Suicidal/Death Ideation, Helplessness/Hopelessness, Self-Doubt, Inefficacy, Stress/Worry, Anxiety, Anger Proneness, Behavior-Restricting Fears, and Multiple Specific Fears. The Externalizing scale contains four individual scales created to identify specific facets of RC4 and RC9 related to acting out behaviors – Juvenile Conduct Problems, Substance Abuse, Aggression, and Activation. The Interpersonal scales contains five individual scales that evaluate for difficulties in interpersonal functioning – Family Problems, Interpersonal Passivity, Social Avoidance, Shyness, and Disaffiliativeness (Friedman et al., 2015).

The two new Interest Scales featured on the MMPI-2-RF were derived from the original clinical scale 5, Masculinity-Femininity (MF). Analyses of clinical scale 5 produced two independent domains: Aesthetic-Literary Interests (AES) and Mechanical-Physical Interests (MEC). These scales can provide information on the respondent's gender-related interests.

The Personality Psychopathology Five (PSY-5) scales were retained from the MMPI-2, however, underwent minor revisions. In developing the adapted version of the PSY-5 scales, Harkness and McNulty (2007) eliminated 22 items used on the original MMPI-2 PSY-5 scales and added 30 new items. The five scales consist of the Aggressiveness-Revised (AGGR-r) scale, a measure of aggressively assertive behavior, the Psychoticism-Revised (PSYC-r) scale, a measure of thought disturbance and disconnection from reality, the Disconstraint-Revised (DISC-r) scale, a measure of

impulsiveness and risk taking behavior, the Negative Emotionality/Neuroticism-Revised (NEGE-r) scale, a measure of a variety of negative emotions including anxiety, insecurity, fear, and worry, and the Introversion/Low Positive Emotionality-Revised (INTR-r) scale, a measure of social disengagement and anhedonia. Descriptions of the MMPI-2-RF scales are provided in Table 3.

(continues)

Table 3

The MMPI-2-RF scales

Scale	Description
Validity Scales	
Cannot Say score	Omitted or double marked responses
Variable Response Inconsistency scale (VRIN-r)	Inconsistent or random responding
True Response Inconsistency scale (TRIN-r)	Response bias or fixed responding
Infrequent Responses (F-r)	Infrequent responses in the general population
Infrequent Psychopathology Responses (Fp-r)	Infrequent responses in psychiatric populations
Infrequent Somatic Responses (Fs-r)	Infrequent somatic complaints in medical patient populations
Symptom Validity (FBS-r)	Non-credible cognitive and somatic complaints
Uncommon Virtues (L-r)	Infrequently claimed moral attributes or activities
Adjustment Validity (K-r)	Uncommonly high level of psychological adjustment
Higher-Order (H-O) scales	
Emotional/Internalizing Dysfunction (EID)	Problems with mood and affect
Thought Dysfunction (THD)	Problems with disordered thinking
Behavioral/External Dysfunction (BXD)	Problems with under-controlled behavior
Restructured Clinical (RC) scales	
Demoralization (RCd)	General distress and emotional discomfort
Somatic Complaints (RC1)	Various physical health complaints
Low Positive Emotions (RC2)	Lack of positive emotionality
Cynicism (RC3)	Beliefs that others are untrustworthy
Antisocial Behavior (RC4)	Rule breaking and imprudent behaviors
Ideas of Persecution (RC6)	Beliefs that others pose a threat
Dysfunctional Negative Emotions (RC7)	Maladaptive anxiety, anger, irritability
Aberrant Experiences (RC8)	Unusual perceptual experiences and disordered thinking
Hypomanic Activation (RC9)	Over-activation, aggression, impulsivity, grandiosity
Specific Problems (SP) scales	
Somatic scales	
Malaise (MLS)	General sense of poor health and physical debilitation
Gastrointestinal Complaints (GIC)	Nausea, upset stomach, poor appetite, vomiting
Head Pain Complaints (HPC)	Head and neck pain
Neurocognitive Complaints (NUC)	Dizziness, numbness, weakness, balance problems
Cognitive Complaints (COG)	Memory and concentration difficulties
Internalizing scales	
Suicidal/Death Ideation (SUI)	Suicidal ideation and recent attempts
Helplessness/Hopelessness (HLP)	Beliefs that problems cannot be solved
Self-Doubt (SFD)	Lacks self-confidence, feelings of uselessness
Inefficacy (NFC)	Beliefs that one is indecisive and ineffective

(cont.)

Table 3 (cont.)

Scale	Description
Stress/Worry (STW)	Preoccupation with disappointments, trouble with stress
Anxiety (AXY)	Pervasive anxiety, frights, sleep disturbances
Anger Proneness (ANP)	Easily angered, impatient
Behavior-Restricting Fears (BRF)	Fears that significantly hinder normal activities
Multiple Specific Fears (MSF)	Specific phobias
Externalizing scales	
Juvenile Conduct Problems (JCP)	Problems at school and home, illegal behavior
Substance Abuse (SUB)	Current or past misuse of drugs or alcohol
Aggression (AGG)	Physically aggressive behaviors
Activation (ACT)	Heightened excitement and energy level
Interpersonal scales	
Family Problems (FML)	Conflictual family relationships
Interpersonal Passivity (IPP)	Unassertive and submissive behaviors
Social Avoidance (SAV)	Avoiding or not enjoying social interactions
Shyness (SHY)	Anxious and inhibited around others
Disaffiliativeness (DSF)	Disliking people and being around them
Interest scales	
Aesthetic-Literary Interests (AES)	Interest in literature, music, theatre
Mechanical-Physical Interests (MEC)	Interest in building things, sports, outdoor activities
Personality Psychopathology Five (PSY-5) scales	
Aggressiveness-Revised (AGGR-r)	Instrumental aggression
Psychoticism-Revised (PSYC-r)	Disconnection from reality
Disconstraint-Revised (DISC-r)	Under-controlled behavior
Negative Emotionality/Neuroticism-Revised (NEGE-r)	Anxiety, worry, fear, insecurity
Introversion/Low Positive Emotionality-Revised (INTR-r)	Anhedonia, social disengagement

Note. Adapted from Ben-Porath and Tellegen (2011) and Friedman et al. (2015)

Assessment of Defensiveness in Personality Testing

Self-report questionnaire measures are susceptible to both unintentional and deliberate response distortions. However, a unique and important contribution that many clinical personality tests provide are validity scales that assess for response distortions (Haywood et al., 1993). For a variety of reasons, respondents completing self-report measures may distort their responses by overreporting their symptoms (i.e., malingering or "faking-bad") or underreporting them (i.e., defensiveness or "faking-good").

Underreporting personal flaws and adjustment difficulties enables respondents to present a positive impression of themselves. Intentional strategies they may use in order to appear favorable and well-adjusted include deliberate denial or minimization of their psychological symptoms, denial of common faults or weaknesses, and attribution of positive and socially desirable characteristics to themselves (Rogers & Bender, 2018). Respondents may also distort their presentation in a self-deceptive manner, providing an overly positive impression of themselves that they believe to be true.

Defensive responding is a relatively common occurrence in both clinical and nonclinical settings. However, evaluators are more likely to encounter this type of response distortion in settings in which evaluees have much to gain from concealing psychological maladjustment (Baer & Miller, 2002). For example, individuals applying for jobs or training programs may be motivated to present themselves in an unrealistically favorable light. Parents undergoing a child custody evaluation may minimize symptoms of psychopathology or deny common shortcomings in an attempt to appear well-adjusted (Baer et al., 1995). A meta-analysis conducted by Baer and Miller (2002) found that an estimated 30% of child custody referrals and job applicants responded defensively on the MMPI-2. Additionally, respondents undergoing forensic evaluations often have significant motivations to present themselves in a particular light. Specifically, sex offenders have been found to frequently deny or minimize psychological problems on personality tests when undergoing clinical evaluations (Grossman & Cavanaugh, 1990; Tarescavage et al., 2018). Given the high-stake outcomes of court proceedings, it is imperative that evaluators interpret results based only on valid data collected from a personality measure. Thus, the accuracy of the validity scales to assess for honest

responding or defensive responding is critical as significant response distortions would undermine the interpretability of the test results and would not provide an accurate depiction of the evaluee's psychological functioning (Rogers & Bender, 2018).

Assessment of Defensiveness on the MMPI, MMPI-2, MMPI-2-RF

The MMPI was one of the first self-report personality measures to include empirically based validity scales designed to detect for response distortions (Baer & Miller, 2002). The traditional validity scales of the original MMPI that detect defensive responding include the Lie scale (L) and the Correction scale (K). The L scale measures a denial of common shortcomings or weaknesses, while the K scale measures a denial of psychopathology. The F-K Index is also commonly used to detect for defensiveness, derived from subtracting the raw scores between the Infrequency scale (F), a measure of psychological disturbance and distress, and the K scale (Baer et al., 1992). Since the MMPI's inception, an abundance of other validity scales have been developed to detect defensive responding. Combinations of the traditional validity scales have been utilized for this purpose and include the L+K Index and the L+K-F Index. Other indices, not based on the traditional MMPI validity scales, include Wiener and Harmon's Subtle-Obvious (S-O) scales, Cofer, Chance, and Judson's Positive Malingering (Mp) scale, Edward's Social Desirability (Esd) scale, Wiggins's Social Desirability (Wsd) scale, and Butcher and Han's Superlative Self-Presentation (S) scale.

Numerous empirical studies have been conducted to evaluate the traditional and supplementary validity scales' ability to detect defensiveness or underreporting of psychopathology on the original MMPI. Baer et al. (1992) conducted a meta-analysis based on 25 MMPI studies in which respondents who answered honestly were compared

to respondents who underreported their psychopathology. Indices examined in this study included the L and K scales, F-K, L+K, social desirability scales (Wsd, Esd), Subtle/Obvious scales (Wo, Ws), and the Positive Malingering scale (Mp). These underreporting indices produced an average effect size of 1.05, suggesting that respondents who answered honestly differed from respondents who underreported their psychopathology by approximately one standard deviation on these validity scales. Average effect sizes for the L and K scales were noted to fall just under one standard deviation of difference between these two groups. The Mp and Wsd scales produced the largest mean effect sizes at approximately 1.5 standard deviations. Despite these supplementary scales' large effect sizes, Baer et al. (1992) advised evaluators to primarily utilize the L and K scales when making judgments regarding defensive responding due to the firm support these scales have gained. In summary, this meta-analysis provides evidence that the MMPI validity indicators designed to detect underreporting are effective in identifying this response style.

The traditional validity scales, L and K, were retained on the MMPI-2 and underwent minimal revisions to their item content. However, many of the supplementary scales used to detect underreporting lost items on their associated scales, suggesting the need for reevaluation of their predictive ability (Baer et al., 1992). The first study that evaluated the effectiveness of the MMPI-2's validity scales to detect a fake-good or fake-bad response style was conducted by Graham and colleagues (1991). Graham et al. (1991) instructed a nonclinical sample of university students (n = 106) to complete the MMPI-2 twice, once with standard instructions and again with instructions to either fake-good or fake-bad. MMPI-2 results were also collected from a clinical sample of

hospitalized psychiatric patients (n = 50) instructed to respond honestly, which was utilized as a comparison to a nonclinical faking-bad sample. The nonclinical sample in the fake-good condition produced mean profiles in which the L and K validity scale scores were much higher than the F scale scores, consistent with an underreporting response style. Additionally, the T-scores on all but one of the clinical scales fell below 50. Comparison of scale score differences between the nonclinical standard instruction condition and the nonclinical faking-good condition showed few significant clinical scale differences. However, there were significant differences in validity scale elevations between these two groups. The fake-good condition produced significantly higher L and K scores and significantly lower F scores compared to the standard instruction condition. The L scale was found to be the best scale at detecting fake-good profiles among men, while the L scale and the L+K Index were equally as effective for women. In summary, the results of this study support that the MMPI-2 validity scales operate in a similar manner to the original MMPI validity scales.

Using the same data collected from the previous study, Timbrook et al. (1993) evaluated the effectiveness of the Wiener-Harmon Subtle-Obvious (S-O) scales (i.e., Subtle scales, Obvious scales, O-S Index), supplementary validity scales developed for the original MMPI, in detecting fake-good and fake-bad responders on the MMPI-2. As the S-O scales were reported to be among the most popular scales used to detect deviant test-taking approaches, Timbrook et al. (1993) recognized the need to assess whether the S-O scales contributed any new information to the traditional validity scales. Consistent with the study's hypothesis, the nonclinical sample instructed to fake-good produced elevations on the L scale. Additionally, the L scale scores were significantly higher than

the L scale scores of the nonclinical sample instructed to fake-bad. Consistent with original MMPI research findings, individuals instructed to fake-good produced significantly higher Subtle scale scores than those instructed to fake-bad. As anticipated, the Obvious scale scores and O-S Index scores were significantly lower for individuals in the fake-good condition. Effect sizes for the L and S-O scales were computed to evaluate the magnitude of group differences between the sample instructed to respond honestly with the sample instructed to fake-good. The L scale produced a large effect size while the Obvious scales, sum of the Subtle scales, and the O-S Index produced small to medium effect sizes. Using a hierarchical regression analysis, Timbrook et al. (1993) found that the S-O scales did not provide significant additional contribution in differentiating between the honest responders and fake-good responders after accounting for the L scale. Therefore, the S-O scales did not provide incremental validity to the already well-established L scale. The results of this study suggest that evaluators should rely on the standard validity scales in identifying fake-good profiles as the S-O scales were not as effective in discriminating between honest and fake-good responding.

Bagby et al. (1994) collected MMPI-2 profiles from a nonclinical sample of university students (n = 244) and a clinical sample of psychiatric inpatients (n = 95) to examine the effectiveness of the standard validity scales (i.e., L, K) and supplementary validity scales (i.e., O-S, F-K, CI, Mp) in detecting fake-good and fake-bad response styles. The scale scores produced by the nonclinical fake-good sample were compared to a nonclinical control group and a clinical sample asked to complete the test under standard instructions. The fake-good group scored significantly lower on 9 of the 10 clinical scales compared to the nonclinical control group and on 8 of the 10 clinical scales

compared to the clinical group. Participants in the fake-good group were successfully able to minimize symptoms of psychopathology, as the mean T-scores on the clinical scales fell consistently in the normal range. Regarding the standard and supplementary validity scales, the fake-good group scored significantly higher on the L, K, and Mp scales and significantly lower on the F-K, O-S, and CI scales than the nonclinical fake-bad group, nonclinical control group, and clinical group, thus demonstrating that these validity scales functioned as designed to detect defensive responding.

In a follow up study, Bagby et al. (1995) evaluated the effectiveness of the MMPI-2's validity scales and supplementary scales in detecting fake-good and fake-bad profiles. The MMPI-2 was administered to a nonclinical sample of university students (n = 344), a clinical sample of general psychiatric patients (n = 129), and a clinical sample of forensic psychiatric patients (n = 159). Participants in the nonclinical sample completed the test under one of three conditions: standard instructions, instructions to fake-bad, or instructions to fake-good. For the purposes of evaluating defensive responding, only participants in the nonclinical standard and fake-good conditions were used. The validity scales under investigation for the fake-good condition included the L scale, S-O scales, and the Mp scale. In evaluating the mean profile differences between the honest responding and fake-good sample, there were significant differences in scores for the L scale, sum of Subtle scales, sum of Obvious scales, and the Mp scale. The fake-good sample produced higher scores on L, Mp, and the sum of Subtle scales and lower scores on the O-S Index and the sum of Obvious scales, consistent with reported findings from Timbrook et al. (1993). A hierarchical regression analysis was conducted to examine the predictive validity of the L scale, O-S Index, and the Mp scale in

differentiating between the fake-good and honest responding samples. Unlike Timbrook et al.'s (1993) findings, the L scale was no more effective in detecting fake-good profiles than the O-S Index when averaged across genders. However, the L scale was found to be a much stronger predictor for women than men. No gender differences were noted for the O-S Index. In examining the Mp scale, this scale was no more or less effective than the L scale when averaged across genders. Similar to the L scale, the Mp scale was found to be a much stronger predictor for women when used on its own. Therefore, the L and Mp scales were better predictors of fake-good responding for women than for men. In summary, Bagby et al. (1995) found mixed results on the predictive ability of the L, S-O, and Mp scales in detecting fake-good profiles across genders and were unable to recommend using one index over another. Despite the O-S Index's more stable predictive ability compared to the L scale and Mp scale, Bagby et al. (1995) cautioned evaluators against relying on this scale to detect for fake-good responding due to its mixed findings in previous studies.

In a similar study, Baer et al. (1995) compared the MMPI-2 profiles of a sample of nonclinical university students instructed to fake-good (n = 50) and a sample of nonclinical university students instructed to respond honestly (n = 50) in order to evaluate the effectiveness of the traditional and supplementary validity scales in detecting defensive responding. Baer et al. (1995) found that the fake-good sample produced significantly lower T-scores on most of the clinical scales compared to the honest responding sample. Additionally, the fake-good sample obtained significantly higher scores on all of the underreporting scales evaluated in this study (i.e., L, K, F-K, L+K, Mp, Wsd, Esd, S, Test-Taking Defensiveness (Tt), Other Deception (Od), Positive

Mental Health (PMH4)) in comparison to the honest responding sample. These underreporting scales produced a mean effect size of 1.8, indicating that the fake-good sample differed from the honest responding sample by nearly 2 standard deviations. Using a hierarchical regression analysis, Baer et al. (1995) found that the Wsd and S scales had significant incremental validity over the L and K scales. The combination of the L, K, Wsd, and S scales were more effective in correctly classifying fake-good responders (90%) than the L and K scales alone (76%). Based on these results, Baer et al. (1995) recommended that evaluators supplement the traditional validity scales, L and K, with the Wsd and S scales in order to improve accuracy in discriminating between honest responding and underreporting profiles.

Bagby et al. (1997) conducted the first study in which a clinical sample was asked to fake-good on the MMPI-2 in order to examine the effectiveness of the standard and supplementary validity scales in detecting defensive responding. A sample of patients diagnosed with schizophrenia (n = 38) and a nonclinical sample of university students (n = 49) completed the MMPI-2 twice, administered one month apart, once with standard instructions and again with instructions to fake-good. Standard validity scales (i.e., L, K) and supplementary validity scales (i.e., F-K, Esd, Tt, L+K Index, PMH4, Mp, Od, S, Wsd) were used to assess for fake-good responding. Under fake-good instructions, the clinical sample scored significantly lower on 8 of the 10 clinical scales and the nonclinical sample scored significantly lower on all of the clinical scales than when both samples responded honestly to the test. Both the clinical and nonclinical samples under fake-good instructions produced higher scores on all of the underreporting validity indicators than when they completed the test under standard instructions. These findings

suggest that clinical patients are capable of modifying how they present themselves on the MMPI-2, and that the validity scales will detect this response distortion. Additional findings indicate that some of the validity scales provide greater predictive power under certain assessment contexts. The Od and S scale were best at detecting fake-good profiles among the nonclinical sample, suggesting these scales be used when assessing nonclinical individuals (e.g., personnel selection). The Esd scale was best at detecting fake-good profiles among the clinical sample, suggesting this scale be used when evaluating individuals previously diagnosed with a mental disorder who may be motivated to minimize or deny their symptoms (e.g., patients seeking to be discharged from an inpatient hospital). In summary, Bagby et al. (1997) found that the traditional validity scales examined were generally not as effective in identifying fake-good profiles among clinical and nonclinical samples compared to the less commonly used supplementary validity scales. However, as this was the first study to examine fake-good responding in a clinical sample, Bagby and colleagues (1997) recommended that evaluators await replication of these results before relying on these supplementary validity scales.

Baer and Miller (2002) conducted a meta-analysis on 14 studies that compared participants given standard instructions to participants instructed to or believed to have underreported on the MMPI-2. Indices examined in this study included traditional validity scales (i.e., L, K) and supplementary validity scales (i.e., F-K, L+K, Mp, Wsd, Esd, Tt, Od, PMH4, S, O-S). These underreporting indices produced an average effect size of 1.25, indicating that participants who underreported differed from participants who responded honestly by more than one standard deviation. This mean effect size is

notably larger than the mean effect size of 1.05 found in Baer et al.'s (1992) meta-analytic review of the original MMPI. The L and K scales produced large mean effect sizes of 1.19 and 1.13, respectively. However, the Wsd scale produced the largest effect size at 1.56. Nonetheless, Baer and Miller (2002) emphasized the importance of utilizing the traditional L and K scales as they have demonstrated a reasonable ability to detect for underreporting and have an extensive body of research supporting their use in comparison to the less frequently studied supplementary validity scales.

The traditional validity scales, L and K, were revised for the MMPI-2-RF, however, they function in a similar manner to their MMPI-2 counterparts (Rogers & Bender, 2018). The Uncommon Virtues scale (L-r) contains items from the MMPI-2's L, Wsd, and Mp scales while the Adjustment Validity scale (K-r) contains items from the MMPI-2's K and S scales. Ben-Porath and Tellegen reported that L-r and K-r are highly correlated with their MMPI-2 scale equivalents. Sellbom and Bagby (2008) conducted the first study to evaluate the predictive ability of these underreporting scales of the MMPI-2-RF. A sample of nonclinical university students completed the MMPI-2-RF under fake-good instructions (n = 65) or standard instructions (n = 67). A comparison sample of individuals who completed the MMPI-2-RF as part of a child custody evaluation (n = 117) represented the suspected underreporting group. The nonclinical sample instructed to fake-good scored significantly lower on 8 of the 9 clinical scales (i.e., RC scales) compared to the nonclinical sample instructed to respond honestly. Similarly, the child custody sample scored significantly lower on 6 of the 9 clinical scales (i.e., RC scales) compared to the honest responding sample. Sellbom and Bagby (2008) found that both the nonclinical fake-good sample and the child custody sample produced

significantly higher mean scores on the L-r and K-r validity scales compared to the nonclinical honest responding sample. As hypothesized, the fake-good sample and child custody sample did not differ significantly on these validity scale scores. The L-r and K-r scales produced mean effect sizes of .89 and .76 in differentiating between the honest responding sample and child custody sample. These results suggest that the L-r and K-r validity scales function as they were designed, to detect underreporting. Results of a hierarchical logistical regression conducted to examine the predictive validity of L-r and K-r in differentiating between underreporting profiles and honest responding profiles were shown to have significant incremental validity to each other, suggesting that both of these scales can be used in conjunction to identify underreporting response styles. This suggests that the L-r and K-r scales are able to differentiate between honest responding and underreporting profiles and function in a similar manner to their original L and K scale counterparts.

Crighton et al. (2017) conducted a study to evaluate the utility of the MMPI-2-RF's underreporting validity scales (i.e., L-r, K-r) under compliant and noncompliant conditions. MMPI-2-RF data was collected from an archival sample of nonclinical university students (n = 302) who were separated into three groups: standard compliant group, underreporting compliant group, and the underreporting noncompliant group. Participants were assigned to a group based on their responses to a post-test questionnaire that assessed their understanding of the test instructions. The standard compliant group was given standard responding instructions, the underreporting compliant group was given instructions to fake-good, and responded in this manner on the test, and the underreporting noncompliant group was given instructions to fake-good, however, denied

doing so on the post-test. As predicted, the underreporting compliant group produced significantly higher scores on the L-r and K-r validity scales compared to the standard compliant group and the underreporting noncompliant group. Differences between the standard compliant group and underreporting compliant group produced large effect sizes on L-r (g = -1.50) and K-r (g = -1.34). Large effect sizes were also observed between the underreporting noncompliant group and underreporting compliant group on L-r (g = -0.96) and K-r (g = -1.28). The underreporting compliant group produced significantly lower RC scale scores across all of the clinical scales compared to both the underreporting noncompliant group and standard compliant group. Crighton et al. (2017) found that L-r and K-r added incrementally to each other in differentiating between underreporting and honest responding profiles. In summary, these results are consistent with those of Sellbom and Bagby's (2008) study, affirming that the MMPI-2-RF's L-r and K-r validity scales are successful in detecting underreporting response styles.

In a similar study, Brown and Sellbom (2020) examined the effectiveness of the MMPI-2-RF's L-r and K-r validity scales in detecting underreporting. A sample of nonclinical university students completed the test under fake-good instructions (n = 236) or under standard instructions (n = 173). The underreporting group produced significantly higher scores on the L-r and K-r validity scales than the standard instruction group. Differences between these two groups produced very large effect sizes on L-r and K-r, with K-r being somewhat larger. The underreporting group scored significantly lower on almost all of the clinical scales than the standard instruction group, with moderate to large effect sizes. Results from a hierarchical logistical regression found that L-r and K-r provided significant incremental validity over one another in differentiating between the

underreporting and standard instruction groups. These results, along with findings from previous studies (Sellbom & Bagby, 2008; Crighton et al., 2017), indicate that L-r and K-r complement one another in predicting underreporting profiles. The results of this study provide further support for the validity and utility of the underreporting scales on the MMPI-2-RF.

Adjusting for Defensive Responding on the MMPI, MMPI-2, MMPI-2-RF

K-correction

In addition to the L and K validity scales that were designed to detect for underreporting, McKinley, Hathaway, and Meehl (1948) developed the K-correction procedure to aid in minimizing the impact of defensive responding on the original MMPI. Since high K scale scores represent defensive responding, McKinley and colleagues developed K-weights that were added to certain clinical scales (i.e., 1, 4, 7, 8, 9) that were shown to be the most susceptible to defensive responding. This procedure was therefore undertaken to adjust for underreporting in an attempt to present a more accurate representation of a respondent's psychopathology. Thus, the goal of the K-correction procedure was to increase the sensitivity of the clinical scales in detecting psychological disturbances. Clinical scales 2, 3, and 6 did not receive K-weights as this procedure resulted in a loss of discriminative power. As scales 5 and 0 are primarily considered to be non-clinical dimensions, they also did not receive K-weights. Although few studies have evaluated the efficacy of the K-correction, this has been a relatively standard practice for most clinicians when interpreting MMPI and MMPI-2 profiles (Friedman et al., 2015).

McKinley, Hathaway, and Meehl originally developed the K scale by evaluating MMPI profiles of 25 male and 25 female psychiatric patients who produced a normal range of clinical scale elevations (T-scores below 70) and showed signs of defensive responding (L scale T-scores above 60). Given these patients' clinical status, they should have produced high clinical scale scores; however, their underreporting response style led to a suppression of clinical scale scores (Dahlstrom et al., 1972). These patient profiles were then compared to a sample of "normals" included in the original MMPI standardization group. An item analysis was conducted that identified 22 MMPI items that discriminated "true and false profiles and their item endorsement by at least 30%" (Dahlstrom et al., 1972, p. 124). High scores on these test items suggest a defensive test taking approach and an increased likelihood of producing a false negative profile, while low scores have a greater likelihood of producing a false positive profile (Friedman et al., 2015). Eight additional items were added to increase the scales' ability to differentiate between the psychiatric sample and the "normals" sample, thus resulting in the 30-item K scale.

Using a new sample of psychiatric patients and "normals," McKinley and colleagues used a differential ratio statistic to determine optimal K-weights, that is, fractions of the K scale raw score, to improve the sensitivity of five of the ten clinical scales in detecting psychopathology. The K-weight is 0.5 for scale 1, 0.4 for scale 4, 1.0 for scale 7 and scale 8, and 0.2 for scale 9. These K-weights are then added to the raw scores of the previously aforementioned clinical scales before conversion to T scores. Thus, the K-weights increase clinical scale scores in an attempt to adjust for

underreporting and reduce the likelihood of obtaining a false negative profile (McKinley et al., 1948).

Since its inception, a number of studies have been conducted to evaluate the usefulness of the K-correction in MMPI and MMPI-2 profiles of clinical and nonclinical samples. Hunt et al. (1948) plotted MMPI profiles of 141 veterans entering a veteran's hospital, both with and without application of the K-correction procedure to assess if the K-correction would increase the diagnostic classification among these psychiatric patients. Three clinical judges unanimously sorted the uncorrected and K-corrected profiles of each patient into one of the following diagnostic categories: psychosis, psychoneurosis, and conduct disorder. Profiles in which the judges could not unanimously agree were excluded from the study. Furthermore, profiles placed in the conduct disorder category were removed from the final analysis due to the infrequency of patients being diagnosed with this disorder in the hospital. The non-K-corrected and K-corrected patient profiles sorted into the psychosis and psychoneurosis categories were then compared to the official hospital diagnosis for each patient. In their first analysis, Hunt et al. (1948) found that only 53 of the 89 cases (58%) plotted using the non-K-corrected profiles were in agreement with the official hospital diagnosis. When the K-corrected profiles were compared to the hospital diagnosis, the diagnostic accuracy increased slightly (61%). This slight increase in diagnostic accuracy with the K-correction applied, however, was statistically non-significant. For the final analysis, researchers removed the small number of conduct-disorder cases and found that the diagnostic accuracy of the psychosis and psychoneurosis categories, both with and without the K-correction, remained ultimately the same as that of the whole sample.

Next, Hunt et al. (1948) evaluated 40 patient profiles that produced borderline elevations (at least one scale above a T-score of 65 but no T-scores above 80) to determine if the K-correction would improve diagnostic accuracy. Similar to the previous findings, the K-correction did not significantly increase diagnostic accuracy although this procedure was specifically designed to be useful for this range of scale scores. Hunt and colleagues (1948) concluded that the K-correction does not provide notable improvement in diagnostic accuracy among veteran psychiatric patients.

Silver and Sines (1962) conducted a study in an attempt to identify whether or not the K-correction or knowledge of the K raw score on the original MMPI would increase the diagnostic accuracy among a sample of psychiatric patients. Without knowing the patients' formal diagnosis upon entry to the hospital, two clinical psychologists "blindly" sorted MMPI profiles of 100 male and 100 female patients into four diagnostic groups (i.e., affective psychotics, schizophrenics, neurotics, personality disorders). Four MMPI profiles were produced for each patient that were later separated into the four diagnostic groups: K scale raw score was not provided but profile was K-corrected, K scale raw score was not provided and profile was not K-corrected, K scale raw score was provided and the profile was K-corrected, and the K scale raw score was provided but the profile was not K-corrected. Results of this study found that neither knowledge of K scale raw scores or knowledge of K-corrected profiles significantly increased diagnostic accuracy.

Wooten (1984) evaluated K-corrected and non-K-corrected MMPI profiles of two samples of Air Force trainees to determine the effectiveness of the K-correction procedure in detecting psychopathology. A sample of 400 trainees identified to have emotional/behavioral problems and a second sample of 200 trainees identified as not

having emotional/behavioral problems completed the MMPI and were scored with and without the K-correction, producing two profiles. A T-score at or above 70 on one or more of the MMPI clinical scores was established as a cutoff. Wooten (1984) found that K-corrected profiles produced an overall hit rate of 80.2%, compared to an overall hit rate of 78.3% for the non-K-corrected profiles. Therefore, both noncorrected and K-corrected profiles were relatively effective in identifying participants who had emotional/behavioral problems from those who did not have problems. The K-correction was found to better identify participants who had emotional/behavioral problems and reduced false negative rates; however, the non-K-correction approach was more effective in identifying individuals without emotional/behavioral problems and reduced false positives rates. Wooten (1984) concluded that the K-correction had a small advantage in identifying participants with psychopathology; however, when the overall hit rate was examined, this advantage became nonexistent due to the relatively high false positive rates with and without the K-correction applied.

Hsu (1986) evaluated the differences in elevations of K-corrected and non-K-corrected MMPI T scores among samples of psychiatric patients (n = 250), nonpatients (n = 640), and medical patients (n = 50,000). The T score distributions for all three samples produced higher clinical scale (i.e., scales 1, 4, 7, 8, 9) scores when the K-correction was applied and lower clinical scale scores when no K-correction was applied. Hsu (1986) applied conventional cut off scores (T = 60 and T = 70) to both K-corrected and non-K-corrected clinical scales to determine false positive and false negative rates among the three samples. When the K-correction procedure was used, false positive rates were typically higher and false negative rates were typically lower as compared to when no K-

correction was applied, for both of the cutoff scores. Hsu (1986) found that K-corrected scores with a cutoff score of T = 60 resulted in a higher proportion of accurate diagnoses among psychiatric patients. In comparison, non-K-corrected scores were found to be more accurate among nonclinical participants. These findings are in accordance with Greene's (1980) recommendation that "clinicians probably need to avoid using K-corrections in settings in which normal persons are being evaluated with the MMPI, but they should use the K-corrections in settings in which psychopathology is suspected" (p. 42).

Colby (1989) evaluated the usefulness of the K-correction in deciphering between nonclinical and clinical patients based on their MMPI profiles. MMPI profiles were obtained from a sample of Caucasian nonclinical patients (n = 289) and a sample of clinical patients (n = 214) and were evaluated with and without the use of the K-correction. The results of this study found that the K-correction reduced false negative rates among the clinical sample, but increased false positives rates among the nonclinical sample. These results confirmed previous findings by Hsu (1986) and Wooten (1984) on the effectiveness of the K-correction in profiles of clinical and nonclinical patients. Due to the mixed findings on the usefulness of the K-correction among various samples, Colby (1989) recommended that evaluators inspect MMPI profiles with and without the K-correction.

McCrae et al. (1989) sought to determine the validity of the K-correction by correlating MMPI clinical scale scores, with and without the K-correction, to an external measure of personality and psychopathology. They administered the MMPI and NEO Personality Inventory (NEO-PI), a self-report measure that assesses five major

dimensions of personality (i.e., neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness), to a sample of psychiatrically normal participants (n = 274). Researchers hypothesized that participants' MMPI clinical scale scores, both with and without the K-correction applied, would be positively correlated to their NEO-PI neuroticism scores and negatively correlated to their agreeableness and conscientiousness scores. Although McCrae et al.'s (1989) hypotheses were supported, they found that the K-correction procedure diminished the overall strength of the correlations between the MMPI clinical scales and the neuroticism, agreeableness, and conscientiousness scales on the NEO-PI. The mean correlation between these three NEO-PI scales and the non-K-corrected MMPI scales was 0.27, and the mean correlation using the K-correction procedure was 0.10. Therefore, the K-correction decreased the validity related to external criteria, suggesting that the K-correction procedure is contraindicated for psychiatrically normal participants.

Despite the little evidence found in support of the K-correction procedure, the traditional K-weights in the original MMPI were retained and remain unchanged in the MMPI-2 (Friedman et al., 2015). Archer et al. (1998) examined the utility of the MMPI-2's K-correction procedure among a large sample of psychiatric inpatients (n = 692). Participants were administered the MMPI-2 and the Symptom Checklist-90-Revised (SCL-90-R), a self-report measure of psychopathology. Additionally, staff psychologists rated each participant using the Brief Psychiatric Rating Scales (BPRS) and the Global Assessment Scale (GAS) to provide an objective measure of the patients' symptoms. Given that previous studies found that the K-correction procedure decreased correlations with external criteria, Archer et al. (1998) hypothesized that the K-corrected clinical scale

scores would have decreased correlations with the SCL-90-R, BPRS, and GAS scores compared to non-K-corrected scores. Their hypothesis was largely confirmed, suggesting that the K-correction procedure is counter-productive as it reduces rather than increases test score validity. These results suggest that the K-correction removes valid variance between test scores rather than removing the impact of defensive responding.

Detrick et al. (2001) evaluated the effectiveness of the K-correction procedure in a sample of police officer candidates (n = 567), a group that often responds defensively on self-report measures. Police officer candidates completed the MMPI-2 and the Inwald Personality Inventory (IPI), a measure of various personality and behavioral characteristics of police officer candidates. Correlations were conducted between K-corrected and non-K-corrected MMPI-2 scales and IPI scales. When the K-correction was not applied, several MMPI-2 clinical scales were "moderately and positively" correlated with IPI scales (Detrick et al., 2001, p. 489). However, when the K-correction was applied, these correlations became negative and significantly decreased in size. Detrick et al. (2001) concluded that non-K-corrected scores should be interpreted with police officer candidates and other normal and highly motivated groups.

Barthlow et al. (2002) sought to determine the effectiveness of the K-correction by correlating MMPI-2 clinical scale scores with and without the K-correction procedure to conceptually relevant external criteria. A sample of 699 outpatients from a community mental health center and 352 clients from a university psychological clinic completed the MMPI-2 and were also rated by their therapists on the Patient Description Form (PDF) or Client Description Form (CDF), rating forms that assess for personality and symptomology. Barthlow and colleagues (2002) rationally matched 1 PDF/CDF scale to

each of the 10 MMPI-2 clinical scales. However, scale 6 and scale 8 were not included in the analysis as there were no significant correlations with a PDF/CDF scale. Correlations between the MMPI-2 clinical scales and matched PDC/CDF scales were higher overall (i.e., 86 of 88 correlations) with non-K-corrected scores. In a majority of cases, there were no significant differences between therapist rating scores and noncorrected and K-corrected clinical scale scores, suggesting little need for the K-correction procedure among relatively nondefensive samples such as this one. A hierarchical logistical regression was conducted with the MMPI-2 K scale scores and clinical scale scores serving as the predictor variables and the therapist ratings serving as the criteria variable to determine optimal K-weights. Results found that the optimal K-weights differed for the clinical samples. Additionally, none of these optimal K-weights were the same as the traditional K-weights, suggesting that other K-weight values may be more accurate in certain settings, particularly with less defensive persons.

Sellbom et al. (2007) evaluated the strength of relationships between K-corrected and non-K-corrected MMPI-2 clinical scale scores with external criterion measures of police officer misconduct. A sample of 291 police officer candidates completed the MMPI-2 as part of their pre-employment screening, were subsequently hired by the police department, and then followed for future police misconduct. Serving as the criterion measure, these officers were evaluated by their supervisor using the Employee Survey (ES), a 26-item questionnaire that inquired about the police officers' problematic behaviors (i.e., excessive force, deceptiveness, abuse of authority). Sellbom et al. (2007) found that overall, non-K-corrected scores had stronger and more significant correlations with their matched ES criteria than the K-corrected scores. Non-K-corrected scores were

found to better predict future problematic behaviors among police officers compared to their K-correction counterpart. Sellbom and colleagues (2007) concluded that the K-correction procedure is counterproductive among relatively defensive samples and therefore recommended that evaluators interpret non-K-corrected scores during pre-employment screenings.

The K-correction has since been discontinued on the MMPI-2-RF. Test developers acknowledged that research found that the K-correction procedure did not increase test score validity and in some instances served to reduce the validity of the MMPI and MMPI-2 clinical scale scores in relation to external criteria. Therefore, the K-r scale on the MMPI-2-RF serves only as a measure of underreporting.

Optimal Cutting Scores

Most psychological measures, including the MMPI and subsequent revisions, have cutting scores that assist evaluators in interpreting test results. Several researchers have proposed optimal cutting scores for detecting an underreporting response style (i.e., defensive responding) on the MMPI, MMPI-2, and MMPI-2-RF; however, research in this area has been limited and inconclusive. Optimal cutting scores ideally maximize the sensitivity and specificity of a test, meaning the test score can accurately identify evaluees as defensive responders when they underreport their symptoms (sensitivity) and accurately exclude evaluees who do not respond defensively (specificity). It is critical that self-report measures accurately identify when evaluees respond defensively as this will impact how evaluators interpret the test results.

Establishing optimal cutting scores that are applicable for all evaluation contexts has been a challenge for researchers. Baer and Miller (2002) noted that optimal cutting

scores vary due to diverse samples of participants across studies. Additionally, some studies emphasize maximizing sensitivity rates over specificity rates and contrariwise, depending on the potential consequences of the evaluation. Optimal cutting scores that maximize sensitivity rates increase the risk of producing false positive results (i.e., misclassifying an honest responder as underreporting) while optimal cutting scores that maximize specificity rates increase the risk of producing false negative results (i.e., misclassifying an underreporting responder as answering honestly). Therefore, no single set of well-established cutting scores have been identified that are effective across all clinical and nonclinical contexts. As a result, some researchers have emphasized that separate optimal cutting scores should be identified for different populations in order to maximize the accurate identification of underreporting on the MMPI (Baer et al., 1992). For example, separate optimal cutting scores should be determined for populations that frequently underreport on self-report measures including parents undergoing custody evaluations, individuals involved in pre-employment evaluations, and alleged sex offenders undergoing evaluations. To date, no study has calculated optimal cutting scores for sex offender populations on the MMPI, MMPI-2, or MMPI-2-RF.

Standard cutting scores for the validity indicators are reported in the MMPI's and subsequent revisions' technical manuals to assist evaluators in test interpretation. According to the MMPI-2-RF manual, L-r T-scores in the 65-69 and 70-79 range reflect potential underreporting; scores in the latter range increase the likelihood that the respondent underreported his or her symptoms. A T-score of 80 or higher likely invalidates the test results as significant underreporting is indicated and the substantive scale scores are no longer interpretable. The interpretative recommendations for the K-r

scale suggest that T-scores in the 60-65 and 66-69 range indicate possible underreporting. Similar to the L-r scale, higher elevations increase the likelihood of underreporting. When T-scores are 70 or higher, the test is likely invalid due to substantial underreporting. Test developers noted that elevations on the L-r and K-r scales may be due to inconsistent responding and thus, VRIN-r and TRIN-r validity scale scores should be evaluated first before concluding that the evaluee was underreporting (Ben-Porath & Tellegen, 2008).

Although optimal cutting scores have been calculated in a multitude of studies examining overreporting on the MMPI and following revisions, there has been a scarcity of research examining optimal cutting scores for underreporting on these instruments (Crighton et al., 2017). A study conducted by Graham et al. (1991), mentioned earlier, assessed MMPI-2 validity scales scores among nonclinical university students. A sample of 106 students completed the MMPI-2 twice, once with standard instructions and a second time with instructions to either fake-good or fake-bad. Optimal cutting scores for the underreporting validity scales (i.e., L scale, K scale, L+K Index, K-F Index) were determined and Graham et al. (1991) concluded that no single cutoff score was found to be effective in accurately identifying men and women in the standard honest responding condition and fake-good condition. The optimal cutoff scores depend on whether it is more important to maximize the accuracy of correctly identifying fake-good respondents or maximize the accuracy of correctly identifying honest responders. Graham et al. (1991) recommended that evaluators use lower cutoff scores when trying to identify a higher percentage of fake-good responders and to use higher cutoff scores when trying to identify higher percentages of honest responders. Of the underreporting scales evaluated,

the K scale best discriminated between the two groups when a raw score cutoff of ≥ 17 was used. This cutoff score correctly identified 76% of fake-good responders and 80% of honest responders.

In a similar study, Austin (1992) administered the MMPI-2 to 110 nonclinical university students and instructed them to either fake-good, fake-bad, or respond honestly to the test. Optimal cutting scores were determined for various validity scales (i.e., L scale, F scale, K scale, O-S scale, F-K Index) using Greene's (1988) proposed cutting scores and were raised or lowered to increase accurate classification rates. Austin (1992) found that the F-K Index, at a raw score cutoff off of < -13, was the best indicator of fake-good respondents, accurately classifying 90% of these respondents. The L scale was second when the T-score cutoff was > 85, accurately classifying 78% of fake-good respondents. The O-S scale was third when the T-score cutoff was < -79, accurately classifying 33% of the respondents. The K scale was fourth when the T-score cutoff was > 69, accurately classifying 30% of respondents. The F scale was last when the raw score cutoff was 0, accurately classifying 5% of respondents. Despite the F-K Index's relative success at identifying fake-good respondents, it incorrectly classified more than one-third of respondents who were instructed to respond honestly. The L scale was less successful in identifying fake-good respondents; however, this scale did not incorrectly classify any honest responders as fake-good responders. The F, K, and O-S scales were not effective in correctly identifying fake-good responders (Austin, 1992).

Bagby et al. (1994) compared a sample of nonclinical university students (*n* = 244) to a clinical sample of psychiatric inpatients (*n* = 95). The nonclinical university students were instructed to either fake-bad, fake-good, or respond honestly to the MMPI-

2 and their validity and clinical scale scores were compared to the clinical patient sample asked to complete the test under standard instructions. In developing optimal cutting scores for the underreporting validity scales (i.e., L scale, K scale, F-K Index, O-S scale, Mp), Bagby et al. (1994) determined that a specificity rate of 80% or higher must be established for the cutting score to be effective. They found that the O-S scale with a cutoff score of < 18 and the F-K Index with a cutoff score of < -12 were the most effective validity scales in detecting defensive responding among the fake-good sample. The overall classification rates of these scales were approximately 80%. Bagby and colleagues (1994) noted that their optimal cutting score for the F-K Index was comparable to Austin's (1992) findings that F-K Index < -13 was the most effective indicator of defensive responding. However, Austin (1992) did not find the O-S scales to be an effective identifier.

Baer et al. (1995) instructed a sample of nonclinical university students to complete the MMPI-2 under fake-good instructions (n = 50) and compared these results to a sample of nonclinical university students who completed the test under standard instructions (n = 50). Baer et al. (1995) found that the Esd scale with a cutting score of ≥ 33 and the PMH4 scale with a cutting score of ≥ 24 were the most effective scales in identifying fake-good profiles. Both scales correctly classified 94% of the fake-good profiles, demonstrating support for these less frequently used research-based measures of underreporting. Regarding the traditional underreporting validity scales, Baer et al. (1995) found that a T-score cutoff of ≥ 57 on the K scale accurately identified 80% of the fake-good profiles, a T-score cutoff of ≥ - 16 on the F-K Index accurately identified 78%

of the profiles, and a T-score cutoff of ≥ 61 on the L scale accurately identified 74% of the profiles.

Crighton et al. (2017) conducted the first study that examined optimal cutting scores for the MMPI-2-RF underreporting scales (i.e., L-r, K-r). The researchers used archival MMPI-2-RF results that were collected from a sample of 302 nonclinical university students who were instructed to either fake-good (i.e., underreporting compliant group) or respond honestly to the test (i.e., standard compliant group). Crighton and colleagues (2017) tested various cutoff scores for L-r (65 T, 70 T, 75 T, and 80 T) and K-r (60 T, 66 T, 69 T, and 72 T) to determine which optimal cutting score best predicted underreporting. When a cutoff score of 65 was applied to the L-r scale, adequate sensitivity was exhibited (.61) and decreased when the cutoff scores were raised. However, the specificity rate was consistently high across all of the cutoff scores (.88-.99). For K-r, adequate sensitivity was exhibited (.54) when a cutoff score of 60 was applied and similarly decreased when the cutoff scores were raised. K-r produced high specificity rates across all of the cutoff scores (.86-1.00). Crighton et al. (2017) determined that the optimal cutoff score for the L-r scale was a T-score of 70, which accurately classified 83% of participants and produced a true positive rate of 56% and a false positive rate of 3%. The optimal cutoff score for the K-r scale was a T-score of 60, which accurately classified 75% of participants and produced a true positive rate of 54% and a false positive rate of 14%.

In a similar study, Brown and Sellbom (2020) sought to determine optimal cutting scores for the L-r and K-r scales on the MMPI-2-RF. They had a sample of nonclinical university students complete the MMPI-2-RF under standard instructions (n = 173) or

instructions to fake-good (n = 236). Brown and Sellbom (2020) tested various cutoff scores for L-r (60 T, 65 T, 70 T, 75 T, and 80 T) and K-r (55 T, 60 T, 65 T, and 70 T) to determine the optimal cutting scores that best detected an underreporting response style. They found L-r displayed high sensitivity of .89 when a cutoff score of 60 was used, and decreased to .25 when the cutoff score was 80. Comparatively, the L-r scale displayed adequate specificity of .74 when a cutoff score of 60 was used, and increased to .99 when the cutoff score was 80. For K-r, adequate sensitivity of .63 was found when a cutoff score of 60 was used, and decreased to .28 when the cutoff score was 70. The K-r scale consistently displayed high specificity across the various cutting score, although a cutoff score of 70 produced the highest specificity rate of .99. Brown and Sellbom (2020) concluded that their results support the respective L-r and K-r cutoff scores of 80 T and 70 T, outlined in the MMPI-2-RF manual.

As evidenced, the focus of most MMPI and subsequent revisions' research had been on developing optimal cutting scores for the validity scales. However, research is needed on developing optimal cutting scores for the clinical scales. Adjusting clinical scale cutoffs for an underreporting response style will improve the interpretability of the test results.

Empirical Findings on MMPI-related Defensiveness in Sex Offenders

As previously mentioned, sex offenders frequently minimize or deny their psychological symptoms on self-report measures while undergoing psychological evaluations. The MMPI and following revisions have been identified as the most commonly used self-report measures used in sex offender evaluations and thus, numerous articles have explored defensive responding patterns among this population.

Early MMPI research on sex offenders primarily focused on comparing participants who denied their engagement in sexually deviant behaviors to participants who admitted to their engagement in sexually deviant behaviors, to determine if these groups produced different response styles on the test. Lanyon and Lutz (1984) conducted one of the first studies that attempted to discriminate between sex offenders who admitted or denied their sex offense using the original MMPI. Ninety indicted or convicted sex offenders were divided into full-denial (n = 18), no-denial (n = 48), or partial-denial (n = 24) groups based on formal police reports. Participants were undergoing a psychological evaluation to assess for rehabilitation potential or insanity/competency to stand trial and thus, had potential motivation to deny or minimize their psychological symptoms on the MMPI. After each participant completed the MMPI, researchers computed the mean validity scale scores, clinical scale scores, and derived validity indices scores (L+K, L+K-F, F-K) across the three groups. Lanyon and Lutz (1984) identified that the partial-denial and full-denial groups responded in a similar manner to the test and therefore, were combined into a single denial group and contrasted with the no-denial group. Compared to the no-denial group, the denial group produced significantly higher scale scores on the L and K scales and on the L+K and L+K-F Indexes and significantly lower scale scores on the F scale and F-K Index, consistent with an underreporting response style. Significant clinical scale score differences between the no-denial and denial group were found on scales 5 (Mf), 6 (Pa), 7 (Pt), 8 (Sc) and 0 (Si). Lanyon and Lutz (1984) found that the L+K-F Index was the best validity indicator at discriminating between the no-denial and denial groups. The researchers concluded that participants who denied their sex offense were more likely to respond in a defensive manner on the MMPI and that the

underreporting validity scales were successful in identifying this type of response distortion.

Grossman and Cavanaugh (1989) administered the MMPI to 36 alleged sex offenders with paraphilic disorders who were undergoing psychological evaluations. The purpose of their study was to determine the degree to which sex offenders will minimize or deny their psychological symptoms. Grossman and Cavanaugh (1989) hypothesized that the alleged sex offenders who denied engaging in paraphilic behaviors would be more likely to minimize their psychopathology than the alleged sex offenders who admitted to these paraphilic behaviors. Additionally, they hypothesized that the alleged sex offenders not facing active legal charges at the time of their psychological evaluation would have a tendency to minimize and show fewer symptoms of psychopathology compared to the alleged sexual offenders who were facing active legal charges. Compared to participants who admitted to paraphilia (n = 20), participants who denied paraphilia (n = 16) had significantly higher scale scores on the K and Mp scales and significantly lower scale scores on the F-K Index, Gough Dissimulation scale (Ds), and O-S Subscales, consistent with an underreporting response style. Additionally, the nonadmitters showed less psychopathology across the 10 clinical scales, with 6 of the 10 scales to a significant degree (Scales 2, 4, 5, 6, 8, and 0). Participants who were not facing legal charges at the time of the evaluation did not show any significant differences in scale scores on the underreporting validity scales compared to the participants who were facing legal chargers or had already been found Not Guilty by Reason of Insanity (NGRI). However, participants who were charged showed less psychopathology on 8 of the 10 MMPI clinical scales than participants who were not charged. Grossman and

Cavanaugh (1989) concluded that many alleged sex offenders experience, yet deny, their psychological symptoms in addition to their sexual disorders. Therefore, evaluators need to maintain alertness to potential "subtle signs of psychopathology" when evaluating sex offenders who deny their paraphilic behaviors (Grossman & Cavanaugh, 1989, p. 885).

Baldwin and Roys (1998) compared MMPI validity and clinical scale scores of alleged sex offenders (n = 114) who admitted to their charges (43%) to offenders who denied their charges (57%). All of the participants were accused of victimizing children and were undergoing a psychological evaluation. The researchers found that the denier group displayed a fake-good style of responding on the MMPI. The denier group differed significantly from the admitter group on the Positive Malingering (Mp) scale after controlling for IQ, demonstrating a defensive response style. The denier group also scored significantly lower than the admitter group on 6 of the 10 clinical scales (Scales 2, 4, 5, 7, 8, and 0) after controlling for IQ. Baldwin and Roys (1998) concluded that the denier group responded more defensively, was less likely to reveal their psychological symptoms, and attempted to deny commonly endorsed everyday problems compared to the admitter group.

Other studies have evaluated defensive responding among sex offenders on the MMPI in conjunction with other personality measures. The descriptions below focus solely on the MMPI results. Wasyliw et al. (1998) examined the MMPI/MMPI-2 and Rorschach Test results of a sample of cleric (n = 33) and noncleric (n = 27) sex offenders who were undergoing psychological evaluations. Wasyliw et al. (1998) first classified participants as minimizers or nonminimizers based on their MMPI/MMPI-2 validity scale scores. Second, the minimizers and nonminimizers were subdivided into two groups,

those who admitted their charges and those who denied their charges. Using the L and K scales and O-S Index as validity indicators of minimization, Wasyliw et al. (1998) applied Greene's (1991) recommended cutoff score of the 75th percentile to separate honest responders from minimizing responders. Applying this cutoff score, 68% of participants showed minimization on the K scale, 70% on the L scale, and 72% on the O-S Index. After dividing the sample into admitters and deniers, the researchers noted that a majority of the deniers were classified as minimizers based on their validity scale scores (81% on K, 76% on L, 90% on O-S Index). As a majority of the sex offenders minimized their psychopathology on the MMPI/MMPI-2, Wasyliw et al. (1998) strongly recommended that evaluators always assess for defensive responding when conducting psychological evaluations with this population.

Grossman et al. (2002) conducted a similar study comparing the MMPI/MMPI-2 and Rorschach Test results of cleric (n = 37) and noncleric (n = 37) sex offenders undergoing a forensic psychological evaluation. Similar to Wasyliw et al.'s (1998) results, a majority of the sex offender sample showed substantial minimizations on the MMPI/MMPI-2 underreporting validity scales (K scale, F-K Index, O-S Index) and were classified as minimizers if they scored above the 75th percentile on one or more of these scales. Participants classified as minimizers produced significantly lower scores across almost all MMPI/MMPI-2 clinical scales compared to the nonminimizers, with 78% of minimizers generating a within-normal-limits profile. Additionally, participants who denied their allegations were more likely to be classified as minimizers on each of the MMPI/MMPI-2 underreporting validity scales compared to participants who admitted their allegations (Grossman et al., 2002).

Haywood et al. (1994) examined MMPI, Multiphasic Sex Inventory (MSI), and Cognition Scale data collected from 59 alleged child molesters undergoing a psychological evaluation to evaluate response distortion patterns among sex offenders using these various measures. Haywood et al. (1994) classified participants as minimizers if their F-K Index on the MMPI was less than or equal to –10 (F-K Index ≤ 10), demonstrating an underreporting response style. Using this cutoff score, 75% of the sample minimized their psychological symptoms. In evaluating the sample's scores across these three measures, Haywood et al. (1994) concluded that a significant percentage of the alleged child molesters who minimized their psychopathology also denied their sexual thoughts, sexual desires, and cognitive distortions.

Other MMPI studies have examined response distortions among various sex offender typologies. Mann et al. (1992) evaluated MMPI-2 validity and clinical scale scores of incarcerated pedophiles in three different sex offender treatment programs: state prison (n = 60), federal prison (n = 24) and a military confinement facility (n = 25). The validity and clinical scale scores of these three samples were combined to provide a generalized normative profile. The researchers found that none of the mean clinical scale scores reached clinical significance (65 T). Scale 4 (Pd) had the highest mean clinical scale score of 63 T. Regarding the underreporting validity scales, the L scale had a mean scale score of 58 T while the K scale had a mean scale score of 51 T. Mann et al. (1992) recommended that evaluators assess for defensive responding when assessing pedophiles. Due to the moderately elevated L scale scores and low K scale scores found in this sample, Mann and colleagues (1992) hypothesized that pedophiles "selectively endorse some problem areas while not endorsing others" (pp. 71).

Another sex offender group that has received research attention are Roman Catholic clergy who have been charged with sexual assault. Plante and Aldridge (2005) examined MMPI-2 profiles of 21 Catholic clergy who were accused of sexual misconduct and were undergoing psychological examinations. All participants in this study admitted their sexual misbehavior. Researchers found that the participants responded in a relatively defensive manner on the test compared to national norms, as evidenced by a mean L scale score of 57.29 and a mean K scale score of 57.38. Twenty-nine percent of participants scored above 65 T on the L scale and 19% scored above 65 T on the K scale, strongly suggesting a defensive response style. Additionally, participants showed significantly low scores across the clinical scales, with no mean clinical scale score reaching clinical significance (65 T).

In a recent study using the MMPI-2-RF, Tarescavage et al. (2018) evaluated MMPI-2-RF and risk assessment data collected from 304 sex offenders who had perpetrated against children and were entering a sex offender treatment program. After examining the mean validity and substantive scale scores, Tarescavage and colleagues (2018) concluded that the sex offender sample demonstrated "relatively high levels of underreporting" compared to the MMPI-2-RF normative sample (pp. 1). Regarding the underreporting validity scales, the mean L-r scale score was 60 T and the mean K-r score was 53 T. Regarding the substantive scales (i.e., H-O scales, RC scales, SP scales, PSY-5 scales), none of the mean scale scores reached clinical significance. The researchers acknowledged that the mean L-r scale score for the offenders was significantly above that of the test's normative sample score whereas the mean K-r scale score was only marginally higher. They considered these differences to mean that sex offenders more

frequently use an overt impression management underreporting response style (measured by L-r) rather than a covert self-deception underreporting response style (measured by K-r). Similar to previous aforementioned studies, Tarascavage et al. (2018) strongly recommended that evaluators assess for underreporting response styles among sex offenders due to their tendency to respond defensively on the MMPI-2-RF. Table 4 shows the mean validity and substantive scale score findings from Tarescavage et al.'s (2018) study.

(continues)

Table 4

Tarescavage et al.'s (2018) MMPI-2-RF sample means and standard deviations for adult male sex offenders

Scale	M	SD
Validity Scales		
Variable Response Inconsistency scale (VRIN-r)	51	11
True Response Inconsistency scale (TRIN-r)	52F	10
Infrequent Responses (F-r)	56	13
Infrequent Psychopathology Responses (Fp-r)	52	12
Infrequent Somatic Responses (Fs-r)	54	13
Symptom Validity (FBS-r)	52	11
Response Bias scale (RBS)	54	11
Uncommon Virtues (L-r)	**60**	12
Adjustment Validity (K-r)	53	11
Higher-Order (H-O) scales		
Emotional/Internalizing Dysfunction (EID)	49	11
Thought Dysfunction (THD)	52	10
Behavioral/External Dysfunction (BXD)	56	10
Restructured Clinical (RC) scales		
Demoralization (RCd)	50	10
Somatic Complaints (RC1)	55	11
Low Positive Emotions (RC2)	50	11
Cynicism (RC3)	51	11
Antisocial Behavior (RC4)	**60**	10
Ideas of Persecution (RC6)	58	12
Dysfunctional Negative Emotions (RC7)	47	10
Aberrant Experiences (RC8)	50	10
Hypomanic Activation (RC9)	46	9
Specific Problems (SP) scales		
Malaise (MLS)	52	11
Gastrointestinal Complaints (GIC)	49	9
Head Pain Complaints (HPC)	52	11
Neurocognitive Complaints (NUC)	57	12
Cognitive Complaints (COG)	51	11
Suicidal/Death Ideation (SUI)	50	11
Helplessness/Hopelessness (HLP)	49	11
Self-Doubt (SFD)	51	11
Inefficacy (NFC)	48	9
Stress/Worry (STW)	49	10
Anxiety (AXY)	50	11
Anger Proneness (ANP)	49	10
Behavior-Restricting Fears (BRF)	49	9
Multiple Specific Fears (MSF)	45	7
Juvenile Conduct Problems (JCP)	**60**	13
Substance Abuse (SUB)	51	9
Aggression (AGG)	48	9
Activation (ACT)	47	11
Family Problems (FML)	48	10

(cont.)

Table 4 (cont.)

Scale	M	SD
Interpersonal Passivity (IPP)	50	10
Social Avoidance (SAV)	55	10
Shyness (SHY)	47	8
Disaffiliativeness (DSF)	50	10
Aesthetic-Literary Interests (AES)	42	8
Mechanical-Physical Interests (MEC)	59	10
Personality Psychopathology Five (PSY-5) scales		
Aggressiveness-Revised (AGGR-r)	51	10
Psychoticism-Revised (PSYC-r)	50	10
Disconstraint-Revised (DISC-r)	57	9
Negative Emotionality/Neuroticism-Revised (NEGE-r)	49	10
Introversion/Low Positive Emotionality-Revised (INTR-r)	54	11

Note. Mean scores in bold are at least one standard deviation above the normative mean.

Of the 42 substantive scales on the MMPI-2-RF, only two of these scales (i.e., JCP, RC4) reached the cut score of 60, falling one standard deviation above the normative mean, and none of these scales reached the clinical cut score of 65. The overwhelming majority of substantive scales fell within the average range.

In conclusion, results from these studies suggest that sex offenders respond in a relatively defensive manner on the MMPI and following revisions. However, individuals who deny their engagement in sex offenses are shown to respond even more defensively on the test compared to offenders who admit to their offenses. These studies have also shown that defensive responding is common across various sex offender typologies. Therefore, this response style should be assessed by evaluators when conducting all evaluations with this population. Patterns that have emerged from these studies indicate that sex offenders who underreport their symptoms frequently produce test profiles in the normative range; however, this is not an accurate portrayal of their psychological functioning.

Chapter 3: Rationale and Purpose of the Study

A compelling amount of evidence across a multitude of MMPI-based studies have found that sex offenders frequently minimize or deny their psychological disturbances and attempt to present themselves in a positive manner. As such, this level of minimization often produces test profiles in which psychological maladjustment is undetected due to a lack of substantive scale score elevations. Evaluators face potential interpretation errors if they equate non-elevated substantive scale scores with psychological adjustment when evaluating sex offenders, especially when the evaluee appears to be responding defensively to the test. Numerous research studies have focused on identifying the utility of the validity indicators in detecting a defensive response style, showing consistent support for the reliability of the traditional validity scales in this form of detection. Although this information is highly useful for evaluators, it does not assist with interpreting the substantive scale scores but only prompts an evaluator to interpret the test results with caution. Unfortunately, there is no standard rubric available to assist evaluators in determining how exactly they should interpret results with caution and to what degree. As a result of this allotted discretion, interpretations may vary significantly from evaluator to evaluator. Due to the high-stakes decisions that follow sex offender evaluations, it is vital that evaluators accurately interpret both the validity and substantive scales on MMPI measures when evaluees respond defensively. Thus, the development of empirically-derived optimal cutting scores for the substantive scale scores will provide a more accurate interpretation of an evaluee's psychological functioning. To date, no previous study has calculated optimal cutting scores for sex offender populations on any version of the MMPI.

The purpose of this study was to develop empirically-derived optimal cutting scores for the MMPI-2-RF substantive scales (i.e., H-O, RC, SP, PSY-5 scales) for sex offender populations using ROC analyses. It was expected that the optimal cutting scores found in this study would be lower than the standard clinical cutting scores reported in the MMPI-2-RF technical manual, demonstrating the need for this study to be conducted.

Chapter 4: Methods

Participants

The participants for this study consisted of two samples. The primary sample consisted of 142 adult men who were under investigation for a sexual offense and had undergone a sex offender evaluation between 2009 and 2018 at a forensic/psychological outpatient treatment facility in Central Florida. This sample had previously been deemed a defensive subgroup through cluster analysis, subsumed under a larger (N = 281) sex offender sample. Inclusion criteria for this sample required that participants be at least 18 years old, had a documented allegation of a sexual offense, and produced a valid MMPI-2-RF profile. MMPI-2-RF inclusion criteria for this sample in the current study consisted of a Cannot Say (CNS) raw score of ≤ 15, demonstrating minimal response omissions, and VRIN-r and TRIN-r scores of ≤ 80, reflecting the absence of significant inconsistent and biased responding. Demographic information and MMPI-2-RF test scores for this sample were obtained from archival data located in the research supervisor's database. The participants in this sample were between the ages of 18 and 73 (M = 38.8, SD = 12.5) The sample was predominantly Caucasian (n = 89; 62.7%), employed (n = 82; 57.7%), and typically reported their highest level of education as higher than a high school diploma (n = 74; 52%). Regarding marital status, participants were typically married (n = 43; 30.3%) or single (n = 42; 29.6%). Table 5 below provides additional, specific demographic information about the sex offender sample, including their offenses and personal and psychological histories.

Table 5

Sex offender sample demographics

Demographic variable	*N*	Percent
Ethnicity		
Caucasian	89	62.7%
Hispanic	27	19%
Black	17	12%
Asian	2	1.4%
Native American	0	0%
Other ethnicity	1	0.7%
Unidentified ethnicity	6	4.2%
Education		
No degree earned	15	10.6%
General Equivalency Diploma	15	10.6%
High school diploma	33	23.2%
Some college education	33	23.2%
Two-year degree	11	7.7%
Four-year degree	21	14.8%
Graduate degree	9	6.3%
Unknown	5	3.5%
Employment Status		
Employed	82	57.7%
Unemployed	17	12%
Unemployed due to arrest	29	20.4%
Disabled	4	2.8%
Retired	3	2.1%
Unknown	7	4.9%
Marital Status		
Single	42	29.6%
Married	43	30.3%
Separated	17	12%
Divorced	30	21.1%
Unknown	10	7%
Living Situation		
Alone	30	21.1%
With significant other	32	22.5%
With roommate	5	3.5%
With parents	20	14.1%
Incarcerated	17	12%
Unknown	38	26.8%
Evaluation Referral Source		
Attorney	72	50.7%
Judge/Court Order	27	19%

(cont.)

Table 5 (cont.)

Demographic variable	*N*	Percent
Case Manager	12	8.5%
Other	28	19.7%
Unknown	3	2.1%
Sex Offense Charge		
Contact offense	45	31.7%
Non-contact offense	28	19.7%
Intended contact	21	14.8%
Allegations without current charge	46	32.4%
Unknown	2	1.4%
Previous Legal History		
None	77	54.2%
Nonviolent	25	17.6%
Violent	11	7.7%
Sex offense	4	2.8%
Combination	20	14.1%
Unknown	5	3.5%
Emotional Abuse History		
Yes	7	4.9%
No	106	74.6%
Unknown	29	20.4%
Physical Abuse History		
Yes	7	4.9%
No	106	74.6%
Unknown	29	20.4%
Sexual Abuse History		
Yes	13	9.2%
No	104	73.2%
Unknown	25	17.6%
Mental Health Treatment History		
Yes	87	61.3%
No	52	36.6%
Unknown	3	2.1%
Sex Offender Treatment History		
Yes	17	12%
No	122	85.9%
Unknown	3	2.1%
Anger Management Treatment History		
Yes	8	5.6%
No	131	92.3%
Unknown	3	2.1%
Substance Abuse Treatment History		
Yes	15	10.6%
No	124	87.3%

(cont.)

Table 5 (cont.)

Demographic variable	*N*	Percent
Unknown	3	2.1%
Substance Use History		
Yes	51	35.9%
No	84	59.2%
Unknown	7	4.9%
Current Substance Use		
Yes	12	8.5%
No	126	88.7%
Unknown	4	2.8%
Mental Health Disorder		
Clinical Disorder		
No diagnosis	43	30.3%
Adjustment disorder	24	16.9%
Mood disorder	12	8.5%
Paraphilia	36	25.4%
Other	23	16.2%
Unknown	4	2.8%
Personality Disorder		
No personality disorder	86	60.6%
Personality disorder	23	16.2%
Personality features	29	20.4%
Unknown	4	2.8%

As seen in Table 5, a substantial portion of the sex offender sample was referred for the evaluation by their attorney (n = 72; 50.7%) and a majority of the sample had been charged with a sex offense (n = 94, 66.2%). Nearly 86% of the sample did not have a history of sex offender treatment, and 73-75% of the sample reported no history of emotional abuse, physical abuse, or sexual abuse. However, 41.6% (n = 59) of the sample had received a clinical diagnosis other than paraphilia and 61.3% (n = 87) of the sample reported having a history of mental health treatment.

The comparison sample consisted of 135 adult men in the community who completed MMPI-2-RF profiles in 2020 and 2021. Inclusion criteria for this sample required that participants be at least 21 years old, reported not having a documented sex offense, were living in Brevard County, Florida or Volusia County, Florida, and produced

a valid MMPI-2-RF profile. MMPI-2-RF inclusion criteria for the comparison sample was the same as the sex offender sample and consisted of a Cannot Say (CNS) raw score of ≤ 15 and VRIN-r and TRIN-r scores of ≤ 80. The participants in this sample were between the ages of 21 and 76 (*M* = 36.8, *SD* = 13.3). The community comparison sample was predominantly Caucasian (*n* = 108; 80%), employed (*n* = 110; 81.5%), and typically reported their highest level of education as higher than a high school diploma (*n* = 132; 97.9%). Table 6 provides additional demographic information about the community comparison sample, including their personal and psychological histories.

(continues)

Table 6

Community comparison sample demographics

Demographic variable	*N*	Percent
Ethnicity		
Caucasian	108	80%
Latinx	11	8.1%
African American	3	2.2%
Asian	3	2.2%
Native American	0	0%
Pacific Islander/Native Hawaiian	0	0%
Mixed/More than one of the above	6	4.4%
Other	3	2.2%
Unknown	1	0.7%
Education		
High school diploma/GED	2	1.5%
Some college education	17	12.6%
Associate's Degree	7	5.2%
Bachelor's Degree	43	31.9%
Some graduate school	14	10.4%
Master's Degree	42	31.1%
Doctoral Degree	9	6.7%
Unknown	1	0.7%
Employment		
Employed	110	81.5%
Unemployed	17	12.6%
Retired	7	5.2%
Unknown	1	0.7%
Marital Status		
Single or never married	59	43.7%
Married	69	51.1%
Divorced or separated	5	3.7%
Widowed	1	0.7%
Unknown	1	0.7%
Legal History		
None	131	97%
DUI/DWI	1	0.7%
Larceny/Theft	1	0.7%
Robbery	0	0%
Sexual offense	0	0%
Aggravated assault	1	0.7%
Domestic violence	1	0.7%
Unknown	0	0%

(cont.)

Table 6 (cont.)

Demographic variable	*N*	Percent
Childhood Physical or Sexual Abuse History		
Yes	11	8.1%
No	123	91.1%
Unknown	1	0.7%
Mental Health Treatment History		
Yes	27	20%
No	106	78.5%
Unknown	2	1.5%
Anger Management Treatment History		
Yes	6	4.4%
No	128	94.8%
Unknown	1	0.7%
Substance Abuse Treatment History		
Yes	4	3%
No	130	96.3%
Unknown	1	0.7%

As seen in Table 6, the community comparison sample is comparable to the sex offender sample in terms of mean age. However, the community comparison sample had a higher percentage of participants with Caucasian ethnicity, as well as higher education and employment status. Nearly all of the community comparison sample denied having a legal history and also reported no history of childhood physical or sexual abuse. In regards to mental health treatment history, only one fifth of the sample reported engaging in these services.

Instruments

MMPI-2-RF

The Minnesota Multiphasic Personality Inventory-Second Edition-Restructured Form (MMPI-2-RF) is the sole instrument to be used in this study. The technical manual for the test provides extensive psychometric findings, supporting that the MMPI-2-RF is a valid and reliable measure of personality and psychopathology (Ben-Porath & Tellegen,

2008). Test data from the MMPI-2-RF normative sample, an outpatient community mental health sample, a psychiatric inpatient hospital sample, and a male psychiatric inpatient sample at a Veteran's Administration hospital were utilized to evaluate the psychometric characteristics of this test. Based on the MMPI-2-RF's normative sample, the test-retest reliability for the validity scales ranged from .40 (TRIN-r) to .84 (K-r) with a Standard Error of Measurement (SEM) that ranged from 4 to 8. Internal consistency reliability for the validity scales were provided for the male subset of the normative sample. For men, the internal consistency reliability ranged from .37 (TRIN-r) to .69 (F-r). The test-retest reliability for the Higher-Order (H-O) scales in the normative sample ranged from .71 (THD) to .91 (BXD) (SEM: 3-5). The internal consistency reliability for the H-O scales ranged from .69 (THD) to .86 (EID) for men in the normative sample. The test-retest reliability for the Restructured Clinical (RC) scales on the normative sample ranged from .64 (RC6) to .89 (RC4) (SEM: 3-6). The internal consistency reliability for the RC scales ranged from .63 (RC6) to .87 (RCd) for men in the normative sample. The test-retest reliability for the Specific Problems (SP) scales including the Somatic/Cognitive, Externalizing, and Internalizing scales, ranged from .54 (NUC) to .88 (SHY) on the normative sample (SEM: 4-7). The internal consistency reliability for the SP scales ranged from .39 (HLP) to .78 (SAV) for men in the normative sample. The test-retest reliability for the Interest scales on the normative sample ranged from .86 (AES) to .92 (MEC) (SEM: 3-4). The internal consistency reliability for the Interest scales ranged from .61 (AES) to .62 (MEC) for men in the normative sample. The test-retest reliability for the Personality Psychopathology Five (PSY-5) scales on the normative sample ranged from .76 (PSYC-r) to .93 (DISC-r) (SEM: 3-5). The internal consistency reliability for

the PSY-5 scales ranged from .69 (PSYC-r) to .77 (INTR-r) for men in the normative sample. Overall, the MMPI-2-RF's test score reliability is deemed to be adequate-to-strong. In testing male sex offenders referred to a sex offender treatment program, Tarescavage et al. (2018) reported that the inter-item correlations, internal consistency, and SEM estimates of the MMPI-2-RF substantive scores generally approximated those for the MMPI-2-RF normative sample.

Test score validity for the MMPI-2-RF's normative sample and a variety of other clinical and forensic samples were also reported in the technical manual (Ben-Porath & Tellegen, 2008). In evaluating the MMPI-2-RF's validity scale functions, the test developers found that the VRIN-r scale on the MMPI-2-RF was comparable to its counterpart on the MMPI-2, and the TRIN-r scale outperformed the MMPI-2 TRIN scale. Both VRIN-r and TRIN-r scales exhibited adequate sensitivity to levels of inconsistent responding. The test developers reported that the F-r, Fp-r, and FBS-r validity scales were largely interchangeable with their MMPI-2 counterparts, and the L-r and K-r scales were effective at detecting the under-reporting of symptoms. In summary, the MMPI-2-RF validity scales perform as well as the original MMPI validity scales and measure what they were designed to measure. Data examining the MMPI-2-RF's construct validity was collected in a wide range of settings where this test is frequently used, including clinical, forensic, medical, and non-clinical settings. Data collected from therapist and administrative staff ratings, clinical diagnoses, record reviews, and other self-report measures were correlated to MMPI-2-RF substantive scales, and demonstrated adequate convergent and discriminant validity. Convergent validity was evaluated by examining the correlations between the MMPI-2-RF substantive scales and comparable MMPI-2

scales, finding that both measures were associated in expected ways. Lastly, intercorrelations between the 50 scales of the MMPI-2-RF were conducted. Among the validity scales, F-r had the highest overall correlations with the substantive scales. In evaluating the intercorrelations between the 42 substantive scales, the H-O scales showed distinctive correlations with the RC, SP, and PSY-5 scales. The intercorrelations of the MMPI-2-RF substantive scales generally correspond to the recommended interpretive structure of the test. Overall, the technical manual demonstrates acceptable-to-strong test score validity, to which further support comes from Tarescavage et al. (2018) in their study with a sex offender population.

Procedure

Data collection began after receiving approval from the Institutional Review Board (IRB) of Florida Institute of Technology (FIT) and the Doctoral Research Project (DRP) Committee. Following approval, the research supervisor made available from her research database the demographic information and MMPI-2-RF test scores of a subsample of sex offenders (n = 142) who were previously evaluated at an outpatient practice site in Orlando, Florida between 2006 and 2018. The sample's test scores were inputted into a confidential Statistical Package for the Social Sciences (SPSS) database and later analyzed in R, a statistical data analysis program.

The comparison sample of community adult men was partially collected (n = 75) for a previous study and supplemented by this student researcher (n = 60). The sample of 135 community adult men residing in Brevard County, Florida and Volusia County, Florida each completed the MMPI-2-RF and a demographic questionnaire. In both data collection processes, adult men in the community were recruited largely by word of

mouth to participate in this study. Additional recruitment procedures included posting fliers in local community centers and spaces (i.e., libraries, Florida Tech campus, apartment complexes), handing out flyers during local community events, and posting on social media pages (i.e., community Facebook groups). Participants who agreed to participate in the study first read and signed the informed consent form via DocuSign before completing the MMPI-2-RF and demographic questionnaire. Due to COVID-19 restrictions related to in-person testing, participants completed the MMPI-2-RF via a Q-Global web-based platform and were monitored over Zoom, an online video communications platform, by the student researchers to ensure participants' compliance with test instructions. The participants' test responses were automatically computer scored in Q-Global and inputted into the SPSS database and R statistical package by the student researchers for analysis. To maintain participants' anonymity, each participant was assigned an identifying number instead of personal identifying information. In the research database, participants were only identified by their identifying number. Their informed consent form containing their name and signature remained separate from the test protocol and demographic data. Following the completion of the study, participants were entered into a drawing to win one of two $50 gift cards to Amazon or Visa.

Data Analyses

Preliminary analyses consisted of generating descriptive statistics to describe both samples. The means and standard deviations of the MMPI-2-RF test scores were also computed separately for both samples. A one-way multivariate analysis of variance (MANOVA) was conducted to determine if there was an overall significant difference between the MMPI-2-RF substantive scale scores between the two samples. The 13

substantive scales that met the assumption of homogeneity were entered into a series of univariate analyses of variance (ANOVAs) to compare scores of the sex offender sample and the community comparison sample. The 27 substantive scales that violated the assumption of equality of error variance were entered into a Mann-Whitney U Test, a nonparametric alternative pairwise comparison method to compare scores between the two samples.

The primary analyses of the study consisted of receiver operating characteristic (ROC) analysis. MMPI-2-RF test scores from the sex offender sample and community sample were entered into an ROC analysis to determine optimal cutting scores for the 40 MMPI-2-RF substantive scales excluding the two interest scales, that is, for the H-O, RC, SP, and PSY-5 scales for the sex offender group. The ROC analysis provided estimates of test-score sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP). Sensitivity and PPP rates were examined to ascertain the probability that a substantive scale score falling above the cutting score will accurately identity the presence of personality maladjustment. Similarly, specificity and NPP rates were examined to ascertain the probability that a substantive scale score falling below the cutting score will accurately identify the absence of personality maladjustment.

Optimal cutting scores for each clinical scale were determined when sensitivity and specificity rates were maximized in order to reduce rates of false positive or false negative classifications. The area under the curve (AUC) was examined to determine the optimal cutting scores' discriminative ability, with higher AUC scores representing higher rates of classification accuracy. As recommended by Streiner and Cairney (2007), an AUC value of .70 was deemed the minimally accepted value, with an AUC value of 1

indicating a perfect classifier. Alternative cutting scores at equal intervals of 40, 45, 50, and 55 were evaluated in terms of sensitivity, specificity, PPP, and NPP to identify the best T score cutoff for practical use.

Chapter 5: Results

Preliminary analyses consisted of deriving means and standard deviations for the MMPI-2-RF scale scores for the sex offender sample and the community comparison sample, shown in Table 7.

Table 7

Means and standard deviations of MMPI-2-RF scale scores for sex offender sample and community comparison sample

	Sex offender sample (N = 142)		Community comparison sample (N = 135)	
Scale	M	SD	M	SD
Validity Scales				
Variable Response Inconsistency scale (VRIN-r)	45.0	7.0	48.4	9.6
True Response Inconsistency scale (TRIN-r)	**55.3**	5.6	**55.6**	5.8
Infrequent Responses (F-r)	46.5	5.8	50.8	10.3
Infrequent Psychopathology Responses (Fp-r)	45.6	6.1	50.0	9.5
Infrequent Somatic Responses (Fs-r)	45.0	5.0	50.9	10.6
Symptom Validity (FBS-r)	49.6	8.7	49.1	10.5
Uncommon Virtues (L-r)	**61.8**	16.1	53.2	9.6
Adjustment Validity (K-r)	**58.8**	6.5	51.9	9.7
Higher-Order (H-O) scales [a]				
Emotional/Internalizing Dysfunction (EID)	42.5	5.9	46.9	10.0
Thought Dysfunction (THD)	47.3	7.1	49.9	9.8
Behavioral/External Dysfunction (BXD)	49.3	8.3	51.0	9.8
Restructured Clinical (RC) scales [a]				
Demoralization (RCd)	43.7	5.4	49.5	10.6
Somatic Complaints (RC1)	47.2	8.4	48.9	10.9
Low Positive Emotions (RC2)	46.3	7.9	46.9	9.9
Cynicism (RC3)	44.8	8.5	50.0	8.8
Antisocial Behavior (RC4)	51.1	8.5	51.8	9.4
Ideas of Persecution (RC6)	51.3	9.6	50.6	9.4
Dysfunctional Negative Emotions (RC7)	40.0	4.8	47.4	9.6
Aberrant Experiences (RC8)	45.3	6.1	51.9	10.5
Hypomanic Activation (RC9)	42.2	6.8	49.3	10.0

(cont.)

Table 7 (cont.)

Scale	Sex offender sample (N = 142) M	SD	Community comparison sample (N = 135) M	SD
Specific Problems (SP) scales				
Somatic scales [a]				
Malaise (MLS)	48.0	8.5	50.0	9.7
Gastrointestinal Complaints (GIC)	47.5	5.1	50.5	9.8
Head Pain Complaints (HPC)	46.5	6.7	49.0	9.4
Neurocognitive Complaints (NUC)	48.1	9.0	51.1	10.5
Cognitive Complaints (COG)	44.6	6.2	52.6	10.4
Internalizing scales [a]				
Suicidal/Death Ideation (SUI)	46.2	5.7	48.3	10.3
Helplessness/Hopelessness (HLP)	44.1	6.8	48.5	9.7
Self-Doubt (SFD)	44.3	4.9	50.2	10.7
Inefficacy (NFC)	44.2	6.6	49.1	10.8
Stress/Worry (STW)	46.3	5.8	51.6	10.5
Anxiety (AXY)	45.3	4.5	48.5	9.1
Anger Proneness (ANP)	42.6	6.4	48.1	9.4
Behavior-Restricting Fears (BRF)	45.4	5.9	47.2	7.6
Multiple Specific Fears (MSF)	45.0	7.3	43.1	6.2
Externalizing scales [a]				
Juvenile Conduct Problems (JCP)	52.6	11.3	49.5	10.7
Substance Abuse (SUB)	48.3	8.7	**55.2**	11.8
Aggression (AGG)	43.1	7.2	48.3	10.2
Activation (ACT)	42.4	7.5	48.1	9.9
Interpersonal scales [a]				
Family Problems (FML)	42.3	6.2	46.3	8.3
Interpersonal Passivity (IPP)	47.3	7.9	48.0	8.4
Social Avoidance (SAV)	50.4	9.4	50.6	9.6
Shyness (SHY)	43.2	7.0	45.9	8.1
Disaffiliativeness (DSF)	47.0	6.8	50.7	10.2
Interest scales				
Aesthetic-Literary Interests (AES)	40.9	7.0	42.7	8.9
Mechanical-Physical Interests (MEC)	**56.1**	9.2	**55.4**	9.1

(cont.)

Table 7 (cont.)

Scale	Sex offender sample (N = 142) M	SD	Community comparison sample (N = 135) M	SD
Personality Psychopathology Five (PSY-5) scales [a]				
Aggressiveness-Revised (AGGR-r)	50.2	7.6	52.3	9.7
Psychoticism-Revised (PSYC-r)	45.9	6.9	50.2	10.4
Disconstraint-Revised (DISC-r)	51.0	8.6	53.9	9.8
Negative Emotionality/Neuroticism-Revised (NEGE-r)	43.8	5.1	48.7	10.8
Introversion/Low Positive Emotionality-Revised (INTR-r)	51.2	9.1	49.8	10.5

Note. [a] = Scales that are the focus of the analyses. Mean scores in bold are at least one-half standard deviation above the normative mean. Mean scores that are underlined are at least one-half standard deviation below the normative mean.

Regarding the sex offender means and standard deviations on the MMPI-2-RF, only 1 (i.e., MEC) of the 42 substantive scales reached a T score of 55, that is, at one-half standard deviation above the normative mean, and none of these scales reached the clinical cut score of 65. Twenty-one substantive scale scores reached a T score of 45, that is, at one-half standard deviation below the normative mean. For the community comparison sample, two of the substantive scales reached a T score of 55 and none of the scales reached the clinical cut score of 65. Three of the substantive scales reached a T score of 45.

A one-way MANOVA was conducted to determine if there was an overall significant difference in MMPI-2-RF substantive scale scores (i.e., H-O, RC, SP, PSY-5 scales) between the sex offender sample and the community comparison sample. The MANOVA results were statistically significant, indicating a significant difference between the sex offender sample and community comparison sample on the substantive scales, Wilk's $\lambda = .479$, $F(40, 236) = 6.419$, $p < .001$, partial $\eta^2 = .521$. In testing for the assumption of equality of covariance matrices using Box's Test, results indicated that the observed covariance matrices of the dependent variables were not equal across groups,

$F(820, 226758.2) = 1.53$, $p < .001$. After examining Levene's Test of Equality of Error Variances for the 40 substantive scales, the 13 scales that met the assumption of homogeneity were entered into a series of univariate analyses of variance (ANOVAs). Six of these substantive scales were found to be statistically significantly different between the two samples. The substantive scales that violated the assumptions of equality of error variance were entered into a Mann-Whitney U Test, a nonparametric alternative pairwise comparison method, to determine if there were statistically significant differences between the two samples. Twenty-three of the 27 substantive scale mean scores that were entered into a Mann-Whitney U Test were significantly different between the sex offender sample and community comparison sample. A total of 29 out of the 40 substantive scales analyzed by either the ANOVA or Mann-Whitney U Test demonstrated a significant difference in scores between the two samples. Results of the ANOVAs and Mann-Whitney U Tests are listed in Tables 8 and 9 respectively.

Table 8

Significant ANOVA results for MMPI-2-RF scale scores of sex offender sample and community comparison sample that met the assumption of equality of covariance matrices

Scale	Sex offender sample (N = 142)		Community comparison sample (N = 135)		F (40,236)	p	η^2
	M	SD	M	SD			
RC3	44.8	8.5	50.0	8.8	25.07	<.001	.08
NUC	48.1	9.0	51.1	10.5	6.25	.013	.02
MSF	45.0	7.3	43.1	6.2	5.08	.025	.02
JCP	52.6	11.3	49.5	10.7	5.55	.019	.02
SHY	43.2	7.0	45.9	8.1	8.75	.003	.03
DISC-r	51.0	8.6	53.9	9.8	7.23	.008	.03

Note. Only scales that reached statistical significance (p <.05) are listed in this table.

Table 9

Significant Mann-Whitney U Test results for MMPI-2-RF scale scores of sex offender sample and community comparison sample

Scale	Sex offender sample (N = 142)		Community comparison sample (N = 135)		U	P
	M	SD	M	SD		
EID	42.5	5.9	46.9	10.0	12229.5	<.001
THD	47.3	7.1	49.9	9.8	10992	.029
RCd	43.7	5.4	49.5	10.6	12637.5	<.001
RC7	40.0	4.8	47.4	9.6	14194	<.001
RC8	45.3	6.1	51.9	10.5	13209	<.001
RC9	42.2	6.8	49.3	10.0	14064	<.001
GIC	47.5	5.1	50.5	9.8	10679.5	.005
HPC	46.5	6.7	49.0	9.4	10775.5	.039
COG	44.6	6.2	52.6	10.4	14023.5	<.001
HLP	44.1	6.8	48.5	9.7	11970	<.001
SFD	44.3	4.9	50.2	10.7	12417	<.001
NFC	44.2	6.6	49.1	10.8	12033.5	<.001
STW	46.3	5.8	51.6	10.5	12601.5	<.001
AXY	45.3	4.5	48.5	9.1	11016	.001
ANP	42.6	6.4	48.1	9.4	13161.5	<.001
BRF	45.4	5.9	47.2	7.6	10638.5	.025
SUB	48.3	8.7	55.2	11.8	12943.5	<.001
AGG	43.1	7.2	48.3	10.2	12509	<.001
ACT	42.4	7.5	48.1	9.9	12767	<.001
FML	42.3	6.2	46.3	8.3	12230.5	<.001
DSF	47.0	6.8	50.7	10.2	11225.5	.001
PSYC-r	45.9	6.9	50.2	10.4	11864	<.001
NEGE-r	43.8	5.1	48.7	10.8	12328.5	<.001

Note. Only scales that reached statistical significance (p <.05) are listed in this table.

The primary analyses of the study consisted of receiver operating characteristic (ROC) analysis to determine empirically-derived optimal cutting scores for MMPI-2-RF substantive scales for sex offenders. Optimal cutting scores were selected to maximize sensitivity and specificity for each of the substantive scales. Table 10 presents the AUC

value, optimal cutting score, sensitivity, specificity, PPP, and NPP values for each substantive scale (i.e., RC, H-O, SP, PSY-5 groups).

Table 10

Optimal cutting scores for MMPI-2-RF substantive scales for the sex offender sample

Scale	AUC	Optimal cutting score	Sensitivity	Specificity	PPP	NPP
EID	.64	48.5	.87	.41	.61	.74
THD	.57	55	.87	.30	.56	.68
BXD	.54	51.5	.66	.44	.55	.55
RCd	.66	50	.87	.45	.63	.77
RC1	.53	57.5	.90	.20	.54	.66
RC2	.50	52	.80	.28	.54	.58
RC3	.68	46.5	.63	.65	.66	.63
RC4	.52	53	.64	.44	.54	.54
RC6	.48	84.5	1	.01	.51	1.0
RC7	.74	47	.93	.50	.66	.87
RC8	.69	54	.92	.40	.62	.83
RC9	.73	45.5	.76	.62	.68	.71
MLS	.57	49	.69	.47	.58	.59
GIC	.56	55	.92	.19	.54	.69
HPC	.56	56	.89	.21	.55	.66
NUC	.58	56	.85	.30	.56	.65
COG	.73	52	.87	.52	.66	.80
SUI	.53	55.5	.95	.11	.53	.68
HLP	.62	46	.70	.51	.60	.62
SFD	.65	47	.80	.45	.61	.69
NFC	.63	52.5	.92	.31	.58	.79
STW	.66	49.5	.75	.53	.63	.67
AXY	.57	51.5	.92	.23	.56	.72
ANP	.69	43	.70	.64	.67	.67
BRF	.55	49.5	.85	.26	.55	.61
MSF	.57	44	.56	.59	.59	.56
JCP	.42	80.5	.99	.01	.51	.50
SUB	.68	52.5	.70	.56	.63	.64
AGG	.65	48	.82	.44	.61	.70
ACT	.67	50.5	.85	.39	.59	.70
FML	.64	46.5	.77	.44	.59	.65
IPP	.52	47.5	.57	.49	.54	.52
SAV	.51	51	.64	.42	.54	.53
SHY	.60	40.5	.45	.71	.62	.55
DSF	.59	51	.81	.34	.56	.63

(cont.)

Table 10 (cont.)

Scale	AUC	Optimal cutting score	Sensitivity	Specificity	PPP	NPP
AGGR-r	.55	62.5	.94	.16	.54	.71
PSYC-r	.62	49.5	.72	.51	.61	.63
DISC-r	.58	52.5	.65	.50	.58	.57
NEGE-r	.64	48	.82	.50	.63	.72
INTR-R	.45	58.5	.81	.22	.52	.53

Note. Optimal cutting scores with decimal points would be rounded up to the nearest whole number for practical use.

The AUC values for RC7 at .74, RC9 at .73, and COG at .73 met the minimum accepted AUC threshold of ≥.70. The AUC values for the remaining scales were below the minimum AUC threshold. Specifically, scales RC6, JCP, and INTR-r produced low AUC values, falling below chance (<.05).

With the exception of the RC6 and JCP scales that produced very low AUC values and exceptionally high optimal cutting scores, the optimal cutting scores for all other scales fell between 40.5 (SHY) and 62.5 (AGGR-r). Specifically, three scales fell between optimal cutting scores of 40-44, 14 scales between 44-49, 13 scales between 50-54, seven scales between 55-59, and one scale between 60-64. Therefore, a majority of the substantive scales produced optimal cutting scores that are close to the MMPI-2-RF's normative mean of 50. Further, all of these optimal cutting scores are below the MMPI-2-RF's clinical cut score of 65.

After examination of the optimal cutting scores, sensitivity, specificity, PPP, and NPP were evaluated for alternative cutting scores of 40, 45, 50, and 55 to provide points of comparison, shown in Table 11.

(continues)

Table 11

Sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP) for MMPI-2-RF substantive scales at alternative cutting scores

Scale	Cutting score	Sensitivity	Specificity	PPP	NPP
EID	≤40	.33	.75	.58	.52
	≤45	.73	.53	.62	.65
	≤50	.89	.33	.58	.73
	≤55	.96	.18	.55	.83
THD	≤40	.35	.67	.52	.49
	≤45	.35	.67	.52	.49
	≤50	.70	.46	.58	.59
	≤55	.87	.30	.56	.68
BXD	≤40	.15	.85	.52	.49
	≤45	.27	.73	.51	.49
	≤50	.66	.44	.55	.55
	≤55	.77	.28	.53	.54
RCd	≤40	.25	.79	.55	.50
	≤45	.56	.63	.61	.57
	≤50	.87	.45	.63	.77
	≤55	.96	.23	.57	.86
RC1	≤40	.21	.78	.50	.48
	≤45	.41	.61	.53	.50
	≤50	.56	.44	.52	.49
	≤55	.83	.27	.54	.60
RC2	≤40	.22	.73	.46	.47
	≤45	.45	.53	.50	.48
	≤50	.80	.28	.54	.58
	≤55	.87	.18	.53	.57
RC3	≤40	.32	.90	.76	.56
	≤45	.53	.75	.69	.60

(cont.)

Table 11 (cont.)

Scale	Cutting score	Sensitivity	Specificity	PPP	NPP
RC3	**≤50**	.77	.43	.59	.64
	≤55	.89	.20	.54	.64
RC4	≤40	.12	.88	.52	.49
	≤45	.20	.76	.46	.47
	≤50	.53	.53	.54	.52
	≤55	.72	.33	.53	.53
RC6	≤40	.00	1.00	--	.49
	≤45	.52	.45	.50	.47
	≤50	.52	.45	.50	.47
	≤55	.52	.45	.50	.47
RC7	≤40	.51	.76	.69	.60
	≤45	.84	.56	.67	.77
	≤50	.98	.34	.61	.94
	≤55	1.00	.19	.56	1.00
RC8	≤40	.42	.75	.64	.55
	≤45	.42	.75	.64	.55
	≤50	.75	.56	.64	.68
	≤55	.92	.40	.62	.83
RC9	≤40	.42	.82	.71	.57
	≤45	.76	.62	.68	.71
	≤50	.89	.38	.60	.76
	≤55	.96	.24	.57	.87
MLS	≤40	.20	.81	.52	.49
	≤45	.20	.81	.52	.49
	≤50	.69	.47	.58	.59
	≤55	.85	.22	.53	.58
GIC	≤40	.00	1.00	--	.49
	≤45	.00	1.00	--	.49
	≤50	.92	.19	.54	.69
	≤55	.92	.19	.54	.69

(cont.)

Table 11 (cont.)

Scale	Cutting score	Sensitivity	Specificity	PPP	NPP
HPC	≤40	.00	1.00	--	.49
	≤45	.66	.43	.55	.55
	≤50	.66	.43	.55	.55
	≤55	.89	.21	.55	.66
NUC	≤40	.00	1.00	--	.49
	≤45	.54	.56	.57	.54
	≤50	.54	.56	.57	.54
	≤55	.85	.30	.56	.65
COG	≤40	.62	.72	.70	.64
	≤45	.62	.72	.70	.64
	≤50	.87	.52	.66	.80
	≤55	.94	.30	.59	.83
SUI	≤40	.00	1.00	--	.49
	≤45	.95	.11	.53	.68
	≤50	.95	.11	.53	.68
	≤55	.95	.11	.53	.68
HLP	≤40	.70	.51	.60	.62
	≤45	.70	.51	.60	.62
	≤50	.70	.51	.60	.62
	≤55	.94	.21	.56	.78
SFD	≤40	.00	1.00	--	.49
	≤45	.80	.45	.61	.69
	≤50	.80	.45	.61	.69
	≤55	.95	.29	.58	.85
NFC	≤40	.28	.79	.58	.51
	≤45	.59	.57	.59	.57
	≤50	.79	.44	.60	.66
	≤55	.97	.22	.57	.88
STW	≤40	.08	.90	.46	.48
	≤45	.47	.73	.64	.57

(cont.)

Table 11 (cont.)

Scale	Cutting score	Sensitivity	Specificity	PPP	NPP
STW	**≤50**	.75	.53	.63	.67
	≤55	.93	.30	.58	.80
AXY	≤40	.00	1.00	--	.49
	≤45	.92	.23	.56	.72
	≤50	.92	.23	.56	.72
	≤55	.92	.23	.56	.72
ANP	≤40	.70	.64	.67	.67
	≤45	.70	.64	.67	.67
	≤50	.88	.39	.60	.75
	≤55	.96	.14	.54	.76
BRF	≤40	.00	1.00	--	.49
	≤45	.85	.26	.55	.61
	≤50	.85	.26	.55	.61
	≤55	.85	.26	.55	.61
MSF	≤40	.27	.68	.48	.47
	≤45	.44	.41	.44	.42
	≤50	.74	.19	.49	.40
	≤55	.96	.01	.51	.25
JCP	≤40	.30	.55	.41	.43
	≤45	.30	.55	.41	.43
	≤50	.60	.32	.48	.43
	≤55	.60	.32	.48	.43
SUB	≤40	.00	1.00	--	.49
	≤45	.49	.75	.67	.58
	≤50	.70	.56	.63	.64
	≤55	.89	.34	.59	.74
AGG	≤40	.48	.72	.64	.57
	≤45	.82	.44	.61	.70
	≤50	.82	.44	.61	.70
	≤55	.89	.26	.56	.69

(cont.)

Table 11 (cont.)

Scale	Cutting score	Sensitivity	Specificity	PPP	NPP
ACT	≤40	.49	.73	.65	.58
	≤45	.73	.50	.61	.64
	≤50	.85	.39	.59	.70
	≤55	.96	.22	.56	.83
FML	≤40	.48	.70	.63	.56
	≤45	.77	.44	.59	.65
	≤50	.92	.24	.56	.74
	≤55	.97	.15	.55	.83
IPP	≤40	.20	.81	.52	.49
	≤45	.37	.65	.53	.49
	≤50	.72	.27	.51	.48
	≤55	.85	.15	.51	.49
SAV	≤40	.11	.90	.54	.49
	≤45	.31	.67	.50	.48
	≤50	.64	.42	.54	.53
	≤55	.80	.21	.52	.51
SHY	≤40	.45	.71	.62	.55
	≤45	.69	.47	.58	.59
	≤50	.89	.18	.53	.60
	≤55	.93	.09	.52	.55
DSF	≤40	.00	1.00	--	.49
	≤45	.81	.34	.56	.63
	≤50	.81	.34	.56	.63
	≤55	.81	.34	.56	.63
AGGR-r	≤40	.08	.93	.55	.49
	≤45	.35	.73	.57	.51
	≤50	.58	.47	.54	.52
	≤55	.73	.36	.54	.56
PSYC-r	≤40	.37	.70	.56	.51
	≤45	.37	.70	.56	.51

(cont.)

Table 11 (cont.)

Scale	Cutting score	Sensitivity	Specificity	PPP	NPP
PSYC-r	**≤50**	.72	.51	.61	.63
	≤55	.86	.35	.58	.70
DISC-r	≤40	.08	.94	.60	.49
	≤45	.27	.83	.63	.52
	≤50	.49	.61	.57	.53
	≤55	.70	.39	.55	.55
NEGE-r	≤40	.30	.76	.57	.51
	≤45	.71	.57	.64	.65
	≤50	.87	.39	.60	.74
	≤55	.99	.21	.57	.93
INTR-r	≤40	.09	.78	.30	.45
	≤45	.30	.61	.44	.45
	≤50	.55	.41	.49	.46
	≤55	.76	.27	.52	.51

Note. -- represents undefined values. Bolded scores indicate the best cutting score based on the optimal cutting score found in Table 10 in addition to the best balance between sensitivity, specificity, PPP, and NPP.

As seen in the table above, a cutting score of ≤55 produced the highest sensitivity rates across all scales and a cutting score of ≤40 produced the highest specificity rates across all scales. A cutting score of ≤50 typically produced the highest PPP values while a cutting score of ≤55 typically produced the highest NPP values. When identifying the cutting score that produced the best balance between sensitivity, specificity, PPP, and NPP, and was in close approximation to the ROC-derived optimal cutting score, a cutting score of ≤40 was optimal for zero scales, ≤45 for 10 scales, ≤50 for 25 scales, and ≤55 for five scales. Figures 1, 2, and 3 found in Appendix C show the final cutting scores across the sets of MMPI-2-RF substantive scales.

Chapter 6: Discussion

Sex offenders undergo psychological evaluations at various stages of the criminal justice proceedings. The findings of these evaluations can be used to determine criminal sentencing. Additionally, these evaluations may extend to identifying the level of dangerousness that sex offenders pose, evaluate their risk for sex offense recidivism, and aid in treatment decisions when indicated. Considering the significant impact the findings of the evaluation may have on offenders' futures, they may be motivated to present themselves in a highly positive manner by minimizing or denying psychological problems in addition to sometimes denying their offense altogether. The MMPI and its revised editions are the most commonly used personality measures in sex offender evaluations, with previous research indicating that sex offenders indeed commonly respond in a defensive manner by underreporting their psychological difficulties as evidenced by elevations on the traditional underreporting validity scales (Haywood et al., 1994; Tarascavage et al., 2018). Despite this common research finding, no previous study has attempted to correct for sex offenders' defensive responding on the substantive scales on any edition of the MMPI. Thus, the purpose of this study was to establish empirically-derived optimal cutting scores for substantive scales of the MMPI-2-RF through comparing scores of a defensive sex offender sample and a community comparison sample. These new cutting scores will assist evaluators in providing a more accurate interpretation of the sex offender's psychological functioning at the time of the evaluation.

The current study began with an examination of the mean MMPI-2-RF scores for a sex offender sample previously identified as producing defensive profiles and

comparing them to those of a community comparison sample of adult men. As is expected from a defensive sample, the defensive subgroup of sex offenders indeed produced most of their mean scores within the average to low range across the MMPI-2-RF substantive scales. In fact, half of their substantive scale score means were at least one-half standard deviation below the normative means. In comparison, only three of the substantive scales were at least one-half standard deviation below the normative means among the community comparison sample. MANOVA results confirmed a significant overall difference in scores between the two samples, with subsequent ANOVA and Mann-Whitney U Tests demonstrating significant differences on 73% of the scales. The community comparison sample produced significantly higher mean substantive scale scores on nearly all of the scales, excepting two, in comparison to the sex offender sample. Stated differently, the sex offender sample produced significantly lower scale scores than the community comparison sample on 68% of the substantive scales. These results, suggesting a noticeable level of underreporting, may appear surprising in the context of a court-ordered evaluation for a sex offense in which honesty might be viewed as favorable. However, considering the high stakes of the evaluation's findings, a tendency toward defensiveness is not unexpected, and defensive denial of personal flaws and maladjustment is indeed reported frequently in studies of sex offenders. The current findings evidence the impact a defensive response style plays in terms of producing lower substantive scale scores, resulting in test profiles that appear to demonstrate psychological adjustment.

This study's descriptive MMPI-2-RF findings for the sex offender sample largely correspond to Tarescavage et al.'s (2018) MMPI-2-RF findings for their sex offender

sample. Across both samples, none of the substantive scale score means reached clinical score levels. However, the defensive sex offender sample in the current study produced somewhat lower mean scores on 39 of the 40 substantive scales of focus than Tarescavage et al.'s (2018) sample, which aligns with their relatively higher mean scores on the L-r scale and K-r scale. The subscale with the highest mean score for both samples (although still within one standard deviation from the normative mean) was Juvenile Conduct Problems (JCP), an externalizing scale that reflects a history of disorderly conduct as an adolescent.

The central focus of this study was to develop optimal cutting scores for H-O, RC, SP, and PSY-5 scales of the MMPI-2-RF for evaluators to use when conducting evaluations with sex offenders in order to adjust for defensive responding. Because these optimal cutting scores were derived from a defensive sex offender sample, they are most appropriate for use in MMPI-2-RF evaluations of sex offenders who produce high L-r and K-r scale scores. ROC analyses revealed that scales RC7, RC9, and COG produced the largest AUC values at .74, .73, and .73. These were the only scales that met the minimally accepted AUC value of .70 recommended by Streiner and Cairney (2007). Thus, the RC7, RC9, and COG scales were best at discriminating between the sex offender sample and community comparison sample. For all three scales, the sex offender sample produced significantly lower mean scores than the community comparison sample, leading to the large differentiation between the two samples.

The RC7 scale in particular provided the best discrimination between the sex offender sample and the community comparison sample, producing the highest AUC value of .74, and with a difference in mean scores being more than a half standard

deviation apart. This finding suggests that clinical symptoms such as anxious rumination and intrusive ideations as measured by RC7 are particularly suppressed by sex offenders' defensive responding. Furthermore, RC7 provided the greatest utility compared to the other substantive scales in terms of sensitivity, specificity, PPP, and NPP. At the optimal cutting score of 47, RC7 produced the highest combined sensitivity (.93) and specificity (.50) rates and the highest combined PPP (.66) and NPP (.87) rates compared to all other substantive scales. The significantly lower scores on RC9 and COG for the sex offender sample represent minimization or denial of impulsivity and cognitive disarray, respectively. It is further noted that RC7 and RC9 are the restructured scale counterparts of MMPI and MMPI-2 clinical scales 7 and 9, both of which were in the set of scales targeted for a K-corrected score to adjust for defensiveness.

With the exception of RC6, JCP, and INTR-r scales that produced AUC values that fell below chance values, optimal cutting scores for the MMPI-2-RF substantive scales ranged from 40.5 and 62.5. These cutting scores were generated statistically to maximize sensitivity and specificity. However, upon further investigation, optimal cutting scores for most of the substantive scales typically prioritized sensitivity over specificity. This is ideal considering a high sensitivity rate places greater emphasis on identifying psychological disturbance in the sex offender sample and therefore produces fewer false negative results. Notably, all of these cutting scores are lower than the traditional cutting score of 65 that is typically used as an indicator of clinical significance on MMPI-2-RF profiles. Substantive scale AGGR-r was the only scale that had an optimal cutting score that fell within the subclinical range of 60-64. A majority of the substantive scales had optimal cutting scores at or below the normative mean of 50. This

demonstrates that the sex offender sample's defensive response style created a large suppression effect on the scale scores, thus demonstrating the need for these new cutting scores to adjust for defensiveness.

Upon further inspection of the ROC-based optimal cutting scores, additional patterns emerge. Substantive scales that encompass internalizing factors of psychopathology typically produced lower optimal cutting scores (e.g., EID, MSF, HLP, ANP) than scales that encompass externalizing factors of psychopathology (e.g., AGG-r, RC4, SUB, ACT). Therefore, internalizing scales are typically more suppressed by defensive responding and require lower cutting scores in order to interpret the scales accurately. Previous research conducted on sex offenders has found that this population typically produces relatively higher scores on externalizing scales in comparison to internalizing scales on the various revisions of the MMPI, consistent with this study's findings (Baldwin & Roys, 1998; Grossman & Cavanaugh, 1990; Lanyon & Lutz, 1984; Mann et al., 1992; Tarascavage et al., 2018).

The ROC-based cutting scores were at T score levels that were not at equal intervals in terms of standard deviations from the MMPI-2-RF normative mean of 50, rendering them difficult to use. Therefore, alternative cutting scores were examined in order to enhance practical applications of the cutting scores in sex offender assessment. Alternative cutting scores were selected at four half standard deviation intervals of 40, 45, 50, and 55, which are the typical cutting score levels selected in defensive responding research. The final selection was guided by consideration of (a) the best balance of sensitivity, specificity, PPP, and NPP, (b) proximity to the ROC-derived optimal cutting score, and (c) the obtained mean score for the sex offender sample. Results indicate that a

cutting score of 50, which is 1.5 standard deviations below the traditional cut score of 65, was optimal for most substantive scales. Specifically, a T score of 45 was optimal for 10 scales (EID, RC7, RC9, HLP, SFD, ANP, ACT, FML, SHY, NEGE-r), a T score of 50 was optimal for 25 scales (THD, BXD, RCd, RC1, RC2, RC3, RC4, RC8, MLS, HPC, NUC, COG, NFC, STW, AXY, BRF, MSF, SUB, AGG, IPP, SAV, DSF, AGGR-r, PSYC-r, INTR-r), and a T score of 55 was deemed optimal for 5 scales (RC6, GIC, SUI, JCP, DISC-r). A T score of 40 was not optimal for any of the substantive scales. These cutting score ranges are comparable to those identified by Tarescavage et al. (2015) for MMPI-2-RF substantive scales, using relative risk ratio analyses to predict problem behavior in a defensive sample of police officers. Specifically, a T score of 45 and a T score of 50 produced reasonable selection ratios for the police officer sample, similar to the current study's findings. Such results suggests that the current study's optimal cutting scores are not unduly low.

Scores above these cutting points may be considered the true point in which some psychological difficulty is acknowledged in the MMPI-2-RF profiles of defensive sex offenders, but possibly at a subclinical level. A more conservative approach might be to consider scores at five T score points above these cutting scores as the level where psychological maladjustment is evident.

The lower MMPI-2-RF cutting scores derived from the current study have practical implications. In using them in sex offender evaluations, evaluators will need to reorient themselves to interpreting relatively low scores as clinically significant when they exceed the new threshold. Additionally, instead of the evaluator using a single cutting score to interpret all of the clinical scales, they will need to consider each

substantive score's individualized cutting score to interpret the scale correctly. However, this is not without precedent as some MMPI-2-RF validity scales, such as VRIN-r and TRIN-r, have cut scores of T 80 rather than T 65, which is familiar to MMPI-2-RF users. Use of these cutting scores could be facilitated by superimposing a "skyline" on the individual evaluee's profile as shown in Figures 1, 2, and 3 (Appendix C), somewhat analogous to the comparison group means depicted in MMPI-2-RF forensic interpretive reports generated by Pearson Assessments.

This study is the first of its kind to develop empirically-derived optimal cutting scores for MMPI-2-RF substantive scales for sex offender populations. Results of this study support that cutting scores for defensive sex offender populations need to be lower than conventional cutting scores in order to accurately interpret their personality traits and psychopathological symptoms. If evaluators rely on traditional cutting scores when interpreting a profile that has high L-r and K-r scores, they will likely not receive an accurate interpretation of the sex offender's functioning due to this population's tendency to underreport symptoms. Additional contributions of this study include expanding the current literature on sex offenders in regards to personality assessment. Prior to this study, only one published empirical study had evaluated sex offenders utilizing the MMPI-2-RF (Tarascavage et al., 2018). Tarascavage et al.'s (2018) study solely evaluated child sex offenders and primarily focused on examining the psychometric properties of this measure. Alternatively, this study included a small subset of sex offenders who perpetrated against adults within a defensive subgroup and included a community comparison sample. Therefore, this study offers a new profile of MMPI-2-RF scores of defensive sex offenders and offers contrasts to MMPI-2-RF scores of adult men

in the community, representative of the general population. These MMPI-2-RF profiles can serve as reference data for future MMPI-related studies. Given the recent release of the MMPI-3 that replicates many of the MMPI-2-RF scales, this study provides important directions that can potentially be extrapolated to this new revision of the test.

An additional strength of this study is that it followed a known-groups design, utilizing a sample of sex offenders that were previously found to produce defensive MMPI-2-RF profiles. Compared to analogue simulation designs that are commonly used in forensic evaluation studies, a known-groups approach is considered an optimal design with the greatest utility (Sellbom et al., 2010). This design has been utilized in recent MMPI-2-RF studies that explore overreporting response styles and malingering in forensic contexts (e.g., Sellbom et al., 2010; Wygant et al., 2009), but to a lesser extent in underreporting and defensiveness studies. Therefore, the current study follows and extends best practices in contemporary research methods.

The limitations of this study must also be considered. Despite having a large enough sex offender sample and community comparison sample to allow for statistical power, larger sample sizes would have allowed for greater generalizability of the results. It should also be noted that the archival sex offender data was obtained from one forensic psychological outpatient center in central Florida and the community sample was collected in a limited geographic area of central Florida. Furthermore, a segment of the sex offender sample had been evaluated more than a decade ago, although this difference in time span is not large enough to represent substantial population shifts. It is also noted that the community sample completed the MMPI-2-RF online in contrast to the paper-and-pencil administration used with the sex offender sample; however, the two forms

have been empirically shown to be equivalent (Finger & Ones, 1999). Future studies should attempt to collect nation-wide samples to increase generalizability. In evaluating the demographic variables between the two samples, it should be noted that the community comparison sample is predominantly Caucasian and achieved a higher level of education than the sex offender sample. More so, the community comparison sample in this study was overall more educated than the general population. Previous research has found that an individual's level of education can greatly impact scores on certain scales of the MMPI-2-RF, specifically on the L and K scales (Friedman et al., 2015). Therefore, because the community comparison sample was highly educated, this may have contributed to the differentiation in MMPI-2-RF scores between the two samples.

Considering the previously aforementioned limitations, replication of this study among a larger national sample of sex offenders that responded defensively on the MMPI-2-RF would support the reliability for the current study's results. It should also be noted that the sex offender sample in the current study was one of three subgroups that was identified through cluster analysis of a larger sex offender sample in a previous study conducted by VanSlyke (2018). A smaller subgroup identified through cluster analysis was a psychological disturbance presentation group (n = 46) in which sex offenders presented with severe emotional disturbance and cognitive difficulties. The current study's cutting scores are not likely to be applicable to them. However, a third subgroup (n = 93) produced within-normal-limits mean scores, which might be the focus of future efforts similar to that undertaken in the current study.

References

Abel, G.G. (1995). *The Abel Assessment for Sexual Interests-2 (AASI-2).* Abel Screening, Inc.

Ahlmeyer, S., Kleinsasser, D., Stoner, J., & Retzlaff, P. (2003). Psychopathology of incarcerated sex offenders. *Journal of Personality Disorders, 17*(4), 306-318. https://doi.org/10.1521/pedi.17.4.306.23969

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

American Psychological Association. (2000). *Personality.* https://www.apa.org/topics/personality#:~:text=Psychology%20Topics%20%2F%2F-,Personality,such%20as%20sociability%20or%20irritability.

American Psychological Association. (2013). *Specialty guidelines for forensic psychology.* https://www.apa.org/practice/guidelines/forensic-psychology

Archer, R.P. (2006). *Forensic uses of clinical assessment instruments.* Lawrence Erlbaum Associates, Inc.

Archer, R.P., Buffington-Vollum, J.K., Stredny, R.V., & Handel, R.W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*(1), 84-94. https://doi.org/10.1207/s15327752jpa8701_07

Archer, R.P., Fontaine, J., & McCrae, R.R. (1998). Effects of two MMPI-2 validity scales on basic scale relations to external criteria. *Journal of Personality Assessment, 70*(1), 87-102. https://doi.org/10.1207/s15327752jpa7001_6

Archer, R.P., Wheeler, E.M., & Vauter, R.A. (2016). Empirically supported forensic assessment. *Clinical Psychology Science and Practice, 23*(4), 348-364. https://doi.org/10.1111/cpsp.12171

Armentrout, J.A., & Hauer, A.L. (1978). MMPIs of rapists of adults, rapists of children, and non-rapist sex offenders. *Journal of Clinical Psychology, 34*(2), 330-332. https://doi.org/10.1002/1097-4679(197804)34:2<330::AID-JCLP2270340213>3.0.CO;2-R

Association for the Treatment of Sexual Abusers. (2005). *Practice standards and guidelines for the evaluation, treatment, and management of adult male sexual abusers.* https://www.atsa.com/atsa-practice-guidelines

Association for the Treatment of Sexual Abusers. (2014). *ATSA practice guidelines for the assessment, treatment, and management of male adult sexual abusers.* https://www.atsa.com/Members/Adult/ATSA_2014_Adult_Practice_Guidelines.pdf

Austin, J.S. (1992). The detection of fake good and fake bad on the MMPI-2. *Education and Psychological Management, 52*(3), 669-674. https://doi.org/10.1177/0013164492052003016

Baer, R.A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment, 14*(1), 16-26. https://doi.org/10.1037/1040-3590.14.1.16

Baer, R.A., Wetter, M.W., & Berry, D.T. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review, 12*(5), 509-525. https://doi.org/10.1016/0272-7358(92)90069-K

Baer, R.A., Wetter, M.W., Nichols, D.S., Greene, R., & Berry, D.T. (1995). Sensitivity of MMPI-2 validity scales to underreporting of symptoms. *Psychological Assessment, 7*(4), 419-423. https://doi.org/10.1037/1040-3590.7.4.419

Bagby, R.M., Buis, T., & Nicholson, R.A. (1995). Related effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment, 7*(1), 84-92. https://doi.org/10.1037/1040-3590.7.1.84

Bagby, R.M., Rogers, R., Buis, T., & Kalemba, V. (1994). Malingered and defensive response styles on the MMPI-2: An examination of validity scales. *Assessment, 1*(1), 31-38. https://doi.org/10.1177/1073191194001001005

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. https://doi.org/10.1037/1040-3590.9.4.406

Baldwin, K. (2015). *Sex offender risk assessment.* U.S. Department of Justice. https://smart.ojp.gov/sites/g/files/xyckuh231/files/media/document/somapi_full_report

Baldwin, K., & Roys, D.T. (1998). Factors associated with denial in a sample of alleged adult sexual offenders. *Sexual Abuse: Journal of Research and Treatment, 10*(3), 211-226. https://doi.org/10.1023/A:1021317920914

Bard, L.A., Carter, D.L., Cerce, D.D., Knight, R.A., Rosenberg, R., & Schneider, B. (1987). A descriptive study of rapists and child molesters: Developmental, clinical, and criminal characteristics. *Behavioral Sciences and the Law, 5*(2), 203-220. https://doi.org/10.1002/bsl.2370050211

Barthlow, D.L., Graham, J.R., Ben-Porath, Y.S., Tellegen, A., & McNulty, J.L. (2002). The appropriateness of the MMPI-2 K correction. *Assessment, 9*(3), 219-229. https://doi.org/10.1177/1073191102009003001

Beech, A.R., Craig, L.A., & Browne, K.D. (2009). *Assessment and treatment of sex offenders: A handbook.* Wiley-Blackwell.

Beech, A.R., Elliott, I.A., Birgden, A., & Findlater, D. (2008). The internet and child sexual offending: A criminological review. *Aggression and Violent Behavior, 13*(3), 216-228. https://doi.org/10.1016/j.avb.2008.03.007

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2 Restructured Form: Manual for administration, scoring, and interpretation.* University of Minnesota Press.

Borum, R., & Grisso, T. (1995). Psychological test use in criminal forensic evaluations. *Professional Psychology: Research and Practice, 26*(5), 465-473. https://doi.org/10.1037/0735-7028.26.5.465

Brown, T.A., & Sellbom, M. (2020). The utility of the MMPI-2-RF validity scales in detecting underreporting. *Journal of Personality Assessment, 102*(1), 66-74. https://doi.org/10.1080/00223891.2018.1539003

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2: Manual for administration and scoring* (Revised Edition). University of Minnesota Press.

Chantry, K., & Craig, R.J. (1994). Psychological screening of sexually violent offenders with the MCMI. *Journal of Clinical Psychology, 50*(3), 430-435. doi:10.1002/1097-4679(199405)50:3 3.0.CO;2-W

Cofer, C.N., Chance, J.E., & Judson, A.J. (1949). A study of malingering on the Minnesota Multiphasic Personality Inventory. *The Journal of Psychology, 27*(2), 491-499. https://doi.org/10.1080/00223980.1949.9917439

Cohen, R. J., & Swerdlik, M. E. (2019). *Psychological testing and assessment: An introduction to tests & measurement* (7th ed.). McGraw-Hill Higher Education.

Colby, F. (1989). Usefulness of the K correction in MMPI profiles of patients and nonpatients. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(2), 142-145. https://doi.org/10.1037/1040-3590.1.2.142

Coric, V., Feuerstein, S., Fortunati, F., Southwick, S., Temporini, H., & Morgan, C.A. (2005). Assessing sex offenders. *Psychiatry, 2*(11), 26-29. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2993520/

Craissati, J. (2005). Sexual violence against women: A psychological approach to the assessment and management of rapists in the community. *Probation Journal, 52*(4), 401-422. https://doi.org/10.1177/0264550505058950

Crighton, A.H., Marek, R.J., Dragon, W.R., & Ben-Porath, Y.S. (2017). Utility of the MMPI-2-RF validity scales in detection of simulated underreporting: Implications of incorporating a manipulation check. *Assessment, 24*(7), 853-864. https://doi.org/10.1177/1073191115627011

Cubellis, M.A., Evans, D.N., & Fera, A.G. (2019). Sex offender stigma: An exploration of vigilantism against sex offenders. *Deviant Behavior, 40*(2), 225-239. https://doi.org/10.1080/01639625.2017.1420459

Cucolo, H.E. & Perlin, M.L. (2013). They're planting stories in the press: The impact of media distortions on sex offender law and policy. *University of Denver Criminal Law Review, 3,* 185-246. https://www.law.du.edu/documents/criminal-law-review/issues/v03-1/planting-stories-in-the-press.pdf

Dahlstrom, W.G., Welsh, G.S., & Dahlstrom, L.E. (1972). *A MMPI handbook, Vol. 1: Clinical interpretation.* University of Minnesota Press.

Davis, K.M., & Archer, R.P. (2010). A critical review of objective-personality inventories with sex offenders. *Journal of Clinical Psychology, 66*(12), 1254-1280. https://doi.org/10.1002/jclp.20722

Derogatis, L.R. (1978). *Derogatis Sexual Functioning Inventory.* Derogatis Measurement Assessments, LLC.

Detrick, P., Chibnall, J.T., & Rosso, M. (2001). Minnesota Multiphasic Personality Inventory – 2 in police officer selection: Normative data and relation to the Inwald Personality Inventory. *Professional Psychology: Research and Practice, 32*(5), 484-490. https://doi.org/10.1037/0735-7028.32.5.484

Drogin, E.Y., Sadoff, R.L., Dattilio, F.M., & Gutheil, T.G. (2011). *Handbook of forensic assessment: Psychological and psychiatric perspectives.* John Wiley & Sons, Inc.

Edens, J.F., Cruise, K.R., & Buffington-Vollum, J.K. (2001). Forensic and correctional applications of the Personality Assessment Inventory. *Behavioral Sciences and the Law, 19*(4), 519-543. https://doi.org/10.1002/bsl.457

Erickson, W.D., Luxenberg, M.G., Walbek, N.H., & Seely, R.K. (1987). Frequency of MMPI two-point code types among sex offenders. *Journal of Consulting and Clinical Psychology, 55*(4), 566-570. https://doi.org/10.1037/0022-006X.55.4.566

Faupel, S., & Przybylski, R. (2015). *Etiology of adult sexual offending.* U.S. Department of Justice. https://smart.ojp.gov/sites/g/files/xyckuh231/files/media/document/

Finger, M.S., & Ones, D.S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment, 11*(1), 58-66. http://dx.doi.org/10.1037/1040-3590.11.1.58

Friedman, A.F., Bolinskey, P.K., Levak, R. W., & Nichols, D.S. (2015). *Psychological assessment with the MMPI-2/MMPI-2-RF* (3rd ed.). Routledge/Taylor & Francis Group.

Graham, J.R. (1987). *The MMPI: A practical guide* (2nd 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*(2), 264-277. https://doi.org/10.1207/s15327752jpa5702_6

Greene, R.L. (1980). *The MMPI: An interpretive manual.* Grune & Stratton, Inc.

Green, R.L. (1988). Assessment of malingering and defensiveness by objective personality measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 123-158). Guilford.

Greene, R.L. (1991). *The MMPI/MMPI-2: An interpretive manual.* Allyn & Bacon.

Greene, R.L. (2000). *The MMPI-2/MMPI-2-RF: An interpretive manual* (3rd ed.). Pearson.

Grossman, S.D., & Amendolance, B. (2017). *Essentials of MCMI-IV assessment.* Wiley.

Grossman, L.S., & Cavanaugh, J.L. (1989). Do sex offenders minimize psychiatric symptoms? *Journal of Forensic Sciences, 34*(4), 881-886. https://doi.org/10.1520/JFS12716J

Grossman, L.S., & Cavanaugh, J.L. (1990). Psychopathy and denial in alleged sex offenders. *The Journal of Nervous and Mental Disease, 178*(12), 739-744. https://doi.org/10.1097/00005053-199012000-00002

Grossman, L.S., Wasyliw, O.E., Benn, A.F., & Gyoerkoe, K.L. (2002). Can sex offenders who minimize on the MMPI conceal psychopathology on the Rorschach? *Journal of Personality Assessment, 78*(3), 484-501. https://doi.org/10.1207/S15327752JPA7803_07

Hall, G.C., Maiuro, R.D., Vitaliano, P.P., & Proctor, W.C. (1986). The utility of the MMPI with men who have sexually assaulted children. *Journal of Consulting and Clinical Psychology, 54*(4), 493-496. https://doi.org/10.1037/0022-006X.54.4.493

Hanson, R.K., & Bussiere, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*(2), 348–362. https://doi.org/10.1037//0022-006x.66.2.348

Hanson, R.K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*(6), 1154–1163. https://doi.org/10.1037/0022-006x.73.6.1154

Harkness, A.R., & McNulty, J.L. (2007). *Restructured version of the MMPI-2 Personality Psychopathology Five (PSY-5) Scale.* [Paper presentation]. American Psychological Association, San Francisco, CA, United States.

Harris, A.J., & Hanson, K. (2010). Clinical, actuarial, and dynamic risk assessment of sexual offenders: Why do things keep changing? *Journal of Sexual Aggression, 16*(3), 296-310. https://doi.org/10.1080/13552600.2010.494772

Harris, R. E., & Lingoes, J. C. (1955). Subscales for the Minnesota Multiphasic Personality Inventory: An aid to profile interpretation. (Mimeographed materials.) San Francisco, CA: University of California, Langley Porter Neuropsychiatric Institute.

Harris, R. E., & Lingoes, J. C. (1968). Subscales for the Minnesota Multiphasic Personality Inventory: An aid to profile interpretation (Rev. ed.). (Mimeographed materials.) San Francisco, CA: University of California, Langley Porter Neuropsychiatric Institute.

Harwood, T.M., Beutler, L.E., & Groth-Marnat, G. (2011). *Integrative assessment of adult personality* (3rd ed.). The Guilford Press.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory.* University of Minnesota Press.

Hayashino, D.S., Wurtele, S.K., & Klebe, K.J. (1995). Child molesters: An examination of cognitive factors. *Journal of Interpersonal Violence, 10*(1), 106-116. https://doi.org/10.1177/088626095010001007

Haywood, T.W., Grossman, L.S., & Hardy, D.W. (1993). Denial and social desirability in clinical evaluations of alleged sex offenders. *Journal of Nervous and Mental Disease, 181*(3), 183-188. https://doi.org/10.1097/00005053-199303000-00006

Haywood, T.W., Grossman, L.S., Kravitz, H.M., & Wasyliw, O.E. (1994). Profiling psychological distortion in alleged child molesters. *Psychological Reports, 75*(2), 915-927. https://doi.org/10.2466/pr0.1994.75.2.915

Hsu, L.M. (1986). Implications of differences in elevations of K-corrected and non-K-corrected MMPI T scores. *Journal of Consulting and Clinical Psychology, 54*(4), 552-557. https://doi.org/10.1037/0022-006X.54.4.552

Hunt, H.F., Carp, A., Cass, W.A., Winder, C.L., & Kantor, R.E. (1948). A study of the differential diagnostic efficacy of the Minnesota Multiphasic Personality Inventory. *Journal of Consulting Psychology, 12*(5), 331-336. https://doi.org/10.1037/h0055841

Koss, M.P., & Butcher, J.N. (1973). A comparison of psychiatric patients' self-report with other sources of clinical information. *Journal of Research in Personality, 7*(3), 225-236. https://doi.org/10.1016/0092-6566(73)90038-X

Krishnamurthy, R., Archer, R.P., & Groth-Marnat, G. (2011). The Rorschach and performance-based assessment. In T.M. Harwood, L.E. Beutler, & G. Groth-Marnat (Eds.), *Integrative assessment of adult personality* (3rd ed., pp. 276-328). Guilford Press.

Lachar, D., & Wrobel, T.A. (1979). Validating clinician's hunches: Construction of a new MMPI critical item set. *Journal of Consulting and Clinical Psychology, 47*(2), 277-284. https://doi.org/10.1037/0022-006X.47.2.277

Langstrom, N., Sjostedt, G., & Grann, M. (2004). Psychiatric disorders and recidivism in sexual offenders. *Sexual Abuse: Journal of Research and Treatment, 16*(2), 139-150. https://doi.org/10.1177/107906320401600204

Lanyon, R.I., & Lutz, R.W. (1984). MMPI discrimination of defensive and nondefensive felony sex offenders. *Journal of Consulting and Clinical Psychology, 52*(5), 841-843. https://doi.org/10.1037/0022-006X.52.5.841

Laulik, S., Allam, J., & Sheridan, L. (2007). An investigation into maladaptive personality functioning in internet sex offenders. *Psychology, Crime, & Law, 13*(5), 523-535. https://doi.org/10.1080/10683160701340577

Laws, D.R., & O'Donohue, W.T. (2008). *Sexual deviance: Theory, assessment, and treatment* (2nd ed.). The Guilford Press.

Levenson, J.S., & Cotter, L.P. (2005). The effect of Megan's Law on sex offender reintegration. *Journal of Contemporary Criminal Justice, 21*(1), 49-66. https://doi.org/10.1177/1043986204271676

Levenson, J.S., & D' Amora, D.A. (2007). Social policies designed to prevent sexual violence: The emperor's new clothes? *Criminal Justice Policy Review, 18*(2), 168-199. https://doi.org/10.1177/0887403406295309

Magaletta, P.R., Faust, E., Bickart, W., & McLearen, A.M. (2012). Exploring clinical and personality characteristics of adult male internet-only child pornography offenders. *International Journal of Offender Therapy and Comparative Criminology, 58*(2), 137-153. https://doi.org/10.1177/0306624X12465271

Maniglio, R. (2012). The role of parent-child bonding, attachment, and interpersonal problems in the development of deviant sexual fantasies in sexual offenders. *Trauma, Violence, & Abuse, 13*(2), 83-96. https://doi.org/10.1177/1524838012440337

Mann, J., Stenning, W., & Borman, C. (1992). The utility of the MMPI-2 with pedophiles. *Journal of Offender Rehabilitation, 18*(3-4), 59-74. https://doi.org/10.1300/J076v18n03_07

Marshall, W.L., & Hall, G.C. (1995). The value of the MMPI in deciding forensic issues in accused sexual offenders. *Sexual Abuse: Journal of Research and Treatment, 7*(3), 205-219. https://doi.org/10.1007/BF02258505

McCabe, M.P., & Wauchope, M. (2005). Behavioural characteristics of rapists. *Journal of Sexual Aggression, 11*(3), 235-247. https://doi.org/10.1080/13552600500272820

McCann, J.T. (1998). Defending the Rorschach in court: An analysis of admissibility using legal and professional standards. *Journal of Personality Assessment, 70*(1), 125-144. https://doi.org/10.1207/s15327752jpa7001_9

McCrae, R.R., Costa, P.T., Dahlstrom, W.G., Barefoot, J.C., Siegler, I.C., & Williams, R.B. (1989). A caution on the use of the MMPI K-correction in research on somatic medicine. *Psychosomatic Medicine, 51*(1), 58-65. https://doi.org/10.1097/00006842-198901000-00006

McGrath, R.J., & Purdy, L.A. (1999). Referring sex offenders for a psychosexual evaluation: A review. *Journal of Addictions and Offender Counseling, 19*(2), 62-75. https://doi.org/10.1002/j.2161-1874.1999.tb00211.x

McKinley, J.C., Hathaway, S.R., & Meehl, P.E. (1948). The Minnesota Multiphasic Personality Inventory: VI. The K scale. *Journal of Consulting Psychology, 12*(1), 20-31. https://doi.org/10.1037/h0061377

Millon, T. (1994). *Manual for the Millon Clinical Multiaxial Inventory (MCMI–III).* National Computer Systems.

Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual.* Psychological Assessment Resources.

Murphy, L., Gottfried, E., DiMario, K., Perkins, D., & Fedoroff, J.P. (2020). Use of penile plethysmography in the court: A review of practices in Canada, the United Kingdom, and the United States. *Behavioral Sciences and the Law, 38*(2), 79-99. https://doi.org/10.1002/bsl.2453

National Center for Missing and Exploited Children. (2016). *Registered sex offenders in the United States and its territories per 100,000 population.* http://www.missingkids. com/en_US/documents/Sex_Offenders_Map.pdf

Nichols, H.R., & Molinder, I. (1984). *Multiphasic Sex Inventory*. Nichols & Molinder Assessments, Inc.

Nichols, H.R., & Molinder, I. (2000). *Multiphasic Sex Inventory, Second edition*. Nichols & Molinder Assessments, Inc.

Phenix, A., & Hoberman, H.M. (2016). *Sexual offending: Predisposing antecedents, assessments and management.* Springer.

Plante, T.G., & Aldridge, A. (2005). Psychological patterns among Roman Catholic clergy accused of sexual misconduct. *Pastoral Psychology, 54*(1), 73-80. https://doi.org/10.1007/s11089-005-6184-8

Polaschek, D.L., Ward, T., & Hudson, S.M. (1997). Rape and rapists: Theory and treatment. *Clinical Psychology Review, 17*(2), 117-144. https://doi.org/10.1016/S0272-7358(96)00048-7

Ramirez, S.R., Jeglic, E., & Calkins, C. (2015). An examination of the relationship between childhood abuse, anger, and violent behavior among a sample of sex offenders. *Health and Justice, 3*(14), 1-10. https://doi.org/10.1186/s40352-015-0025-3

Reijnen, L., Bulten, E., & Nijman, H. (2009). Demographic and personality characteristics of internet child pornography downloaders in comparison to other offenders. *Journal of Child Sexual Abuse, 18*(6), 611-622. https://doi.org/10.1080/10538710903317232

Rogers R., & Bender, S.D. (2018). *Clinical assessment of malingering and deception* (4th ed.). The Guilford Press.

Rouse, S.V. (2017). Test review of the Millon Clinical Multiaxial Inventory-IV. In J.F. Carlson, K.F. Geisinger, & J.L. Jonson (Eds.), *The twentieth mental measurement yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Sachsenmaier, S.J., & Gress, C.L. (2009). In D. Thornton & D.R. Laws, *Cognitive approaches to the assessment of sexual interests in sexual offenders* (pp. 31-57). Wiley-Blackwell.

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. https://doi.org/10.1037/a0012952

Sellbom, M., Fischler, G.L., & Ben-Porath, Y.S. (2007). Identifying MMPI-2 predictors of police officer integrity and misconduct. *Criminal Justice and Behavior, 34*(8), 985-1004. https://doi.org/10.1177/0093854807301224

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. https://doi.org/10.1037/a0018222

Seto, M.C., Hanson, R.K., & Babchishin, K.M. (2011). Contact sexual offending by men with online sexual offenses. *Sexual Abuse: Journal of Research and Treatment, 23*(1), 124-145. https://doi.org/10.1177/1079063210369013

Silver, R.J., & Sines, L.K. (1962). Diagnostic efficacy of the MMPI with and without the K correction. *Journal of Clinical Psychology, 18*(3), 312-314. https://doi.org/10.1002/1097-4679(196207)18:3<312::AID-JCLP2270180321>3.0.CO;2-8

Simons, D.A. (2015). *Sex offender typologies.* U.S. Department of Justice. https://smart.ojp.gov/sites/g/files/xyckuh231/files/media/document/somapi_full_report

Streiner, D.L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry, 52*(2), 121-128. https://doi.org/10.1177/070674370705200210

Tarescavage, A.M., Cappo, B.M., & Ben-Porath, Y.S. (2018). Assessment of sex offenders with the Minnesota Multiphasic Personality/Inventory-2-Restructured Form. *Sexual Abuse: Journal of Research and Treatment, 30*(4), 413-437. https://doi.org/10.1177/1079063216667921

Tarescavage, A.M., Corey, D.M., & Ben-Porath, Y.S. (2015). Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) Predictors of Police Officer Problem Behavior. *Assessment, 22*(1), 116-132. https://doi.org/10.1177/1073191114534885

Thorpe, G.L., & Dawson, R.F. (2010). Test review of the Personality Assessment Inventory. In R.A. Spies, J.F. Carlson, & K.F. Geisinger (Eds.), *The eighteenth mental measurement yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Timbrook, R.E., Graham, J.R., Keiller, S.W., & Watts, D. (1993). Comparison of the Wiener-Harmon Subtle-Obvious scales and the standard validity scales in detecting valid and invalid MMPI-2 profiles. *Psychological Assessment, 5*(1), 53-61. https://doi.org/10.1037/1040-3590.5.1.53

Tomak, S., Weschler, F.S., Ghahramanlou-Holloway, Virden, T., & Nademin, M.E. (2009). An empirical study of the personality characteristics of internet sex offenders. *Journal of Sexual Aggression, 15*(2), 139-148. https://doi.org/10.1080/13552600902823063

United States Department of Justice (n.d.). *National sex offender public website.* https://www.nsopw.gov/en/SafetyAndEducation/QuestionsAndAnswers

VanSlyke, R. (2018). *Profile patterns of sex offenders using the Minnesota multiphasic personality inventory-second edition-restructured form (MMPI-2-RF)* (Unpublished doctoral dissertation). Florida Institute of Technology, Melbourne, Florida.

Wasyliw, O.E., Benn, A.F., Grossman, L.S., & Haywood, T.W. (1998). Detection of minimization of psychopathology on the Rorschach in cleric and noncleric alleged sex offenders. *Assessment, 5*(4), 389-397. https://doi.org/10.1177/107319119800500408

Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of internet child pornography offenders: A comparison with child molesters. *Sexual Abuse: Journal of Research and Treatment, 19*(4), 449-465. https://doi.org/10.1177/107906320701900408

Weiner, I.B. (2003). *Principles of Rorschach interpretation* (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Weiner, I.B., Exner, J.E., & Sciara, A. (1996). Is the Rorschach welcome in the courtroom? *Journal of Personality Assessment, 67*(2), 422-424. https://doi.org/10.1207/s15327752jpa6702_15

Weiner, I.B., & Greene, R.L. (2008). *Handbook of personality assessment.* John Wiley & Sons, Inc.

Wendling, M. (2020, August 20). QAnon: What is it and where did it come from? *BBC News.* https://www.bbc.com/news/53498434

Whitaker, D.J., Le, B., Hanson, R.K., Baker, C.K., McMahon, P.M., Ryan, G., Klein, A., & Rice, D.D. (2008). Risk factors for the perpetration of child sexual abuse: A review and meta-analysis. *Child Abuse & Neglect, 32*(5), 529-548. https://doi.org/10.1016/j.chiabu.2007.08.005

Wiggins, J.S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23*(5), 419-427. https://doi.org/10.1037/h0047823

Wiggins, J. S. (1966). Substantive dimensions of self-report in the MMPI item pool. Psychological Monographs, 80(22, Whole No. 630).

Wiseman, J. (2015). *Incidence and prevalence of sexual offending (part 1).* U.S. Department of Justice.https://smart.ojp.gov/sites/g/files/xyckuh231/files/media/document/incidenceandp revalenceofsexualoffending.pdf

Wood, R.M., Grossman, L.S., & Fichtner, C.G. (2000). Psychological assessment, treatment, and outcome with sex offenders. *Behavioral Sciences and the Law, 18*(1), 23-41. https://doi.org/10.1002/(SICI)1099-0798(200001/02)18:1 3.0.CO;2-D

Wooten, A.J. (1984). Effectiveness of the K correction in the detection of psychopathology and its impact on profile height and configuration among young adult men. *Journal of Consulting and Clinical Psychology*, *52*(3), 468-473. https://doi.org/10.1037/0022-006X.52.3.468

Wygant, D.B., Ben-Porath, Y.S., Arbisi, P.A., Berry, D.T., 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. https://doi.org/10.1093/arclin/acp073

Young, M.H., Justice, J., & Erdberg, P. (2012). A comparison of rape and molest offenders in prison psychiatric treatment. *International Journal of Offender Therapy and Comparative Criminology*, *56*(7), 1103-1123. https://doi.org/10.1177/0306624X11417361

Zachar, P. (2017). Test review of the Millon Clinical Multiaxial Inventory-IV. In J.F. Carlson, K.F. Geisinger, & J.L. Jonson (Eds.), *The twentieth mental measurement yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Zonana, H., Roth, J.A., & Coric, V. (2004). Forensic assessment of sex offenders. In R.I. Simon & L.H. Gold, *The American psychiatric publishing textbook of forensic psychiatry* (pp. 349-375). American Psychiatric Association Publishing.

Appendix A

Informed Consent Form

Please read this consent document carefully before you decide to participate in this study. The researcher will answer any questions you have before you sign this form.

Purpose: This research study is being conducted by Katie Glauner, a clinical psychology doctoral student, under the direction of Dr. Radhika Krishnamurthy at Florida Institute of Technology. You are being asked to participate in this study comparing personality characteristics in adult men using a clinical assessment tool – the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF). Your data will be compared to data of other adult men as part of this study.

Procedures: Upon agreeing to participate in this study, you will be asked to complete a short demographic questionnaire. Next, you will complete the MMPI-2-RF, a self-report personality questionnaire. The total time of participation in this study will be approximately 35-50 minutes.

Voluntary Participation and Potential Risks: Participation in this study is completely voluntary and you have the right to withdraw at any time. There are no expected risks in participating in this study.

Benefits: Your participation in this study will contribute to the research on personality characteristics of groups of adult men and contribute to the literature on the MMPI-2-RF.

Confidentiality: All of your response records and data sources will be assigned a participant identification number to replace your personal identifying information in order to maintain your anonymity and confidentiality in this study. Your name will not be used in any part of this study. This informed consent form, which requires your signature, will be stored separately from all other data sources to ensure confidentiality.

How Data Will Be Used: The results of this study will solely be used for research purposes. Participants will not receive individual feedback regarding their test results and these results will not be shared with anyone else.

Contact Information: If you have any questions or concerns regarding this study, please contact Katie Glauner at kglauner2018@my.fit.edu or Dr. Radhika Krishnamurthy at rkrishna@fit.edu. You may also contact, Dr. Jignya Patel, Florida Tech's IRB Chairperson, if you have any concerns about this study.

Dr. Jignya Patel, IRB Chairperson
150 West University Blvd. Melbourne, FL 32901
Email: jpatel@fit.edu Phone: 321.674.7391

Participant's Agreement: By signing below, I am acknowledging that I have read the above information and am voluntarily agreeing to participate in this study. I am also affirming that I am 18+ years of age.

Participant: _____ Date: _____

Principal Investigator: _____ Date: _____

Appendix B

Demographic Questionnaire

Participant ID: _____

1. Your current age: _____ years

2. Your identified gender:

   Male Female Other

3. Your identified ethnicity:

   African American
   Asian
   Latinx
   Native American
   Pacific Islander/Native Hawaiian
   White/Caucasian
   Mixed/More than one of the above
   Other

4. Your current marital status:

   Single or Never Married
   Married
   Divorced or Separated
   Widowed

5. How many children you have (if any):

   _____ children _____ I do not have any children

6. The highest level of education you completed:

   High school diploma/GED
   Some college
   Associate's Degree
   Bachelor's Degree
   Some graduate school

Master's Degree

7. Your current employment status:

Unemployed

Employed Your current job title: _____

8. Have you ever been charged with any of the following?

DUI/DWI	___ Yes	___ No
Larceny/Theft	___ Yes	___ No
Robbery	___ Yes	___ No
Sexual Offense	___ Yes	___ No
Aggravated Assault	___ Yes	___ No
Domestic Violence	___ Yes	___ No

9. Have you ever experienced childhood physical or sexual abuse?

Yes No

10. Have you ever received treatment for substance abuse?

Yes No

11. Have you ever received treatment for anger management?

Yes No

12. Have you ever received treatment for mental health reasons?

Yes No

Appendix C

Comparisons of Alternative Cutting Scores for Defensive Sex Offender Sample with Traditional Cutting Scores

Figure 1. Higher-Order and Restructured Clinical Scales

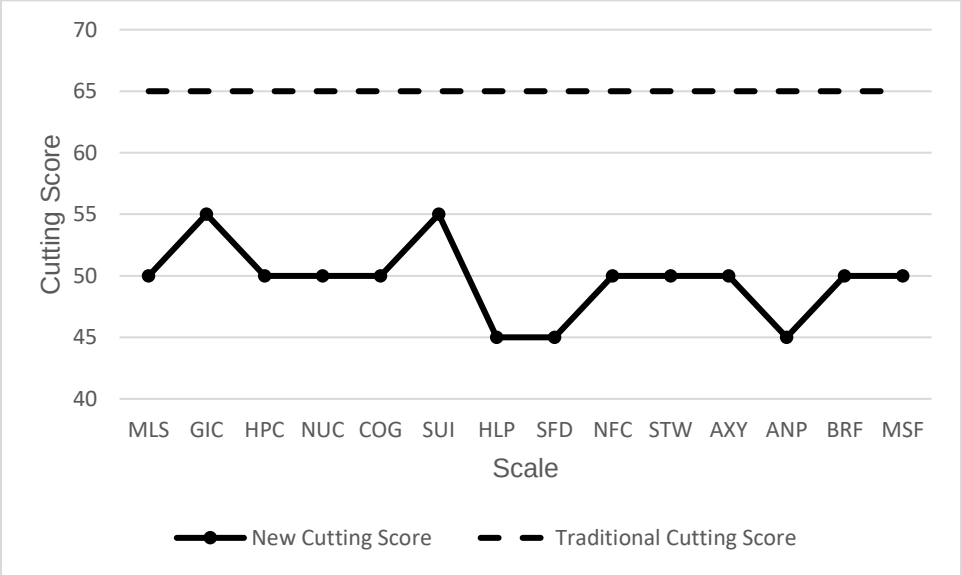

Figure 2. Somatic/Cognitive and Internalizing Scales

Figure 3. Externalizing, Interpersonal, and Personality Psychopathology Five Scales